

Economia Matemática

Aplicações e História

Jorge Paulo de Araújo



ACTUAL



Economia Matemática

Aplicações
e História

Jorge Paulo de Araújo



Economia Matemática

Economia Matemática

Aplicações e História

2022

Jorge Paulo de Araújo



ECONOMIA MATEMÁTICA

APLICAÇÕES E HISTÓRIA

© Almedina, 2022

AUTOR: Jorge Paulo de Araújo

DIRETOR DA ALMEDINA BRASIL: Rodrigo Mentz

EDITOR DE CIÊNCIAS SOCIAIS E HUMANAS: Marco Pace

ASSISTENTES EDITORIAIS: Isabela Leite e Larissa Nogueira

ESTAGIÁRIA DE PRODUÇÃO: Laura Roberti

REVISÃO: Gabriela Leite

DIAGRAMAÇÃO: Almedina

DESIGN DE CAPA: Roberta Bassanetto

CONVERSÃO PARA EBOOK: Cumbuca Studio

ISBN: 9786587019352

Abril, 2022

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Araújo, Jorge Paulo de

Economia matemática : aplicações e história /
Jorge Paulo de Araújo. -- 1. ed. -- São Paulo :
Almedina, 2022.

Bibliografia.

ISBN 978-65-87019-34-5

1. Matemática - Estudo e ensino I. Título.
22-98308 - CDD-510.7

Índices para catálogo sistemático:

1. Matemática : Estudo e ensino 510.7

Aline Grazielle Benitez - Bibliotecária - CRB-1/3129

Este livro segue as regras do novo Acordo Ortográfico da Língua Portuguesa (1990).

Todos os direitos reservados. Nenhuma parte deste livro, protegido por copyright, pode ser reproduzida, armazenada ou transmitida de alguma forma ou por algum meio, seja eletrônico ou mecânico, inclusive fotocópia, gravação ou qualquer sistema de armazenagem de informações, sem a permissão expressa e por escrito da editora.

EDITORA: Almedina Brasil

Rua José Maria Lisboa, 860, Conj.131 e 132, Jardim Paulista | 01423-001 São Paulo | Brasil

www.almedina.com.br

In memoriam dedico este livro ao Hugo e à Elfrides.

AGRADECIMENTOS

Nestes dois anos de pandemia, nos quais este livro foi finalizado, sou grato aos que conheço e aos que não conheço e que providenciaram os meios necessários para a sobrevivência neste trecho das nossas vidas. Entretanto, esta é a ocasião de agradecer àqueles que estiveram mais próximos nestes anos sombrios.

No círculo familiar agradeço aos muitos amados Bruno, Paula, Rosa e Sandra. Um agradecimento especial à Rosa, sem ela não haveria livro. No círculo dos amigos, agradeço ao Atilano, à Denise à Lidiane e à Luíza. Ao Rafa, por me ensinar uma maneira de viabilizar os gráficos.

Cada um ao seu modo, concorreu para humanizar o hábitat no qual vivi enquanto se estendeu este monólogo interior. Sem estas pessoas, nada seria possível. No entanto, tenho certeza de que agradecimentos não compensam uma ínfima parte do apoio e do imprescindível afeto que recebi.

Agradeço aos meus alunos que ao longo de muitos semestres testemunharam a elaboração deste livro. A maneira exata e feliz de expressar a dívida dos professores para com seus estudantes é do Thomas Kuhn no Prefácio do “Estrutura das Revoluções Científicas”: “[...] devo agradecer aos meus alunos pelas lições inestimáveis, tanto acerca da viabilidade de minhas concepções, como a respeito das técnicas apropriadas a sua comunicação eficaz.” Faço minha esta frase iluminada e sobretudo verdadeira.

Por fim agradeço a toda a equipe da Editora Almedina por acreditarem e viabilizarem este projeto, especialmente ao Sr. Marco Pace, pelo entusiasmo com que recebeu a proposta e pela paciência com a qual aguardou sua concretização e a Srta. Gabriela Leite pela meticulosa e exaustiva revisão.

Jorge Araújo

SUMÁRIO

Capa

Folha de Rosto

Página de Créditos

Sumário

Introdução

Lição UMA – Programação Linear

Lição DOIS – A Teoria dos Jogos de Soma-Zero

Lição TRÊS – O Método Simplex

Lição QUATRO – Superfícies e Lagrangeanos

Lição CINCO – O Teorema de Kuhn-Tucker

Lição SEIS – Convexidade

Lição SETE – Condições de 2ª Ordem

Lição OITO – Envelopes e Estática Comparativa

Lição ZERO – Apêndice – Equações Lineares

Introdução

Cientistas e filósofos com inclinações científicas impressionaram-se muito com a precisão dessas disciplinas [matemática e física]. Sentiram-se obrigados a acompanhar ou a estimular essa “exatidão”, esperando provavelmente que a fertilidade surgisse como uma espécie de subproduto da precisão. Todavia, a fertilidade não é decorrência da exatidão, mas da percepção de novos problemas onde ninguém os havia visto antes e da invenção de novas maneiras de resolvê-los.¹

O objetivo deste livro é apresentar os conteúdos básicos que um economista deve saber sobre otimização estática, destacando, através da utilização das fontes primárias nesse assunto, a influência entre as duas disciplinas: Economia e Matemática.

Da primeira até a última lição, o tema central compreende um período de tempo extremamente curto: um ano apenas! Do final de junho de 1949 até o começo de agosto de 1950. Sem dúvida, esse ano excepcional dependeu de progressos anteriores e teve avanços posteriores. Suas consequências prolongam-se até os dias atuais. Esse é o tema do livro.

Este curso se resume à análise de um único resultado, o Teorema de Kuhn-Tucker. Ainda que o Teorema de Kuhn-Tucker não seja a realização histórica de certas ideias, ele nos permite uma visão unificada dos resultados do livro: é como se os progressos particulares conspirassem para se reunirem em um resultado apenas: o Teorema de Kuhn-Tucker.

O livro foi escrito com o objetivo de destacar a cooperação entre duas ciências e a criatividade em ambas. Espero que os leitores descubram a “enorme flexibilidade na formação de conceitos”² (von

Neumann) do modo matemático de pensar, e que isso facilite o acesso para algumas das grandes contribuições científicas do século 20 ao pensamento econômico.

Para muitos, a Economia Matemática não é propriamente uma disciplina científica, mas um conjunto de resultados em teoria econômica transcritos para a linguagem matemática, quando isso é factível. Outros concedem um papel mais ativo para a Matemática, na medida em que os resultados disponíveis necessitam de adaptação para o uso adequado na Economia.

Sem dúvida, às vezes, a Matemática é uma expressão conveniente e inúmeras vezes os resultados precisam ser ajustados para tratar problemas específicos. Portanto, coerentemente, cursos designados por “Economia Matemática” são concebidos como o suporte matemático para a teoria econômica.

Entretanto, economistas e matemáticos desenvolveram matemática original com a finalidade de resolver problemas oriundos da economia, mas esse fato não é tão manifesto nem tão enfatizado quanto os anteriores, principalmente em relação às contribuições de economistas para a matemática.

Em 1932, von Neumann apresentou seu modelo de crescimento³ numa palestra em Princeton. Nessa publicação, ele ofereceu uma versão do Teorema do Ponto Fixo de Brouwer, conhecida como “Teorema da Interseção”, desenvolvida especialmente para aplicação em seu modelo econômico. Esse artigo é, na opinião de muitos, o mais importante da Economia Matemática do século 20:

O período contemporâneo de desenvolvimento da economia matemática está profundamente influenciado por John von Neumann através do seu artigo de 1928 sobre jogos e do seu artigo de 1937 sobre crescimento econômico, que também se destacam como grandes acidentes numa carreira com muitas facetas.⁴

Em 1928, para muitos o ano zero da Economia Matemática, von Neumann publicou o artigo “*Zur Theorie der Gesellschaftsspiele*”⁵. A matemática empregada nas demonstrações não tem nada de original. Os argumentos são aqueles pertencentes ao que chamamos, atualmente, de “Análise na Reta”. De fato, original é a maneira difícil e criativa através da qual von Neumann demonstra a existência de selas para os *payoffs* esperados dos jogadores. Todavia, desse artigo emergiu uma maneira original de pensar a interação dos agentes econômicos e novos conceitos econômicos e matemáticos, tais como equilíbrios em selas, coalisões e quase-convexidade, quase-concavidade.

No começo do século 20, o melhor resultado sobre a definição de formas quadráticas em subespaços estava no livro “*Theory of Maxima and Minima*”,⁶ de 1917, que é a referência de autores como Hotelling. Os resultados como conhecemos atualmente, como no artigo de Gerard Debreu⁷, desenvolveram-se ao longo de duas décadas, motivados principalmente pelo esforço de Hotelling em explorar as condições de 2ª ordem com o objetivo de aprimorar o conhecimento sobre as curvas de demanda e oferta. Nesse caso, o interesse de Hotelling em teoria econômica, conduziu-o a novos resultados em matemática.

O interesse que o Teorema de Kuhn-Tucker despertou entre os economistas motivou muitos aperfeiçoamentos da técnica empregada pelos autores. Destacamos Arrow e seus colaboradores, que analisaram as chamadas condições de qualificação⁸ e generalizaram o resultado de Kuhn e Tucker para funções quase-côncavas e quase-convexas⁹ – outro episódio de contribuição da Economia para a Matemática.

Muito mais conhecida é a Programação Linear, uma disciplina matemática criada para tratar questões originadas de problemas

econômicos. Em 1949, economistas, matemáticos e físicos como Oskar Morgenstern, Paul Samuelson, Robert Dorfman, George Dantzig, Harold Kuhn, Albert Tucker, David Gale e Tjalling Koopmans, reunidos no primeiro congresso de Programação Linear, deram início a uma colaboração em que os conhecimentos especializados não disputaram a superioridade dos métodos, mas se associaram numa construção multidisciplinar para a ciência econômica. Os esforços de von Neumann, Dantzig, Kuhn, Tucker e Gale, tendo em vista os problemas propostos, conduziram a novos resultados em Álgebra das Matrizes e na Teoria de Máximos e Mínimos – que até então não estava muito além do ponto em que Lagrange havia deixado –, como o famoso Teorema de Kuhn-Tucker e sua posterior extensão por Arrow e Enthoven. São resultados matemáticos novos suscitados pela Economia.

Na Teoria Econômica e na Matemática, como nas demais ciências, a pluralidade de métodos e as influências externas sempre foram a causa de inumeráveis avanços, ainda que acompanhados de debates acirrados e posições obstinadas.

Mais recentemente, existe uma noção mais ou menos difundida de que o mérito das formulações matemáticas está na generalidade, na precisão e no rigor lógico, como em Gerard Debreu¹⁰. Essa perspectiva “colonizadora” não faz justiça para ambas as ciências. Como escreveu Leontief, a preocupação com os aspectos formais na Economia “nos leva a acreditar que a prevalência do pensamento ilógico ou pelo menos descuidado constitui o aspecto mais vulnerável de nossa disciplina em seu estado atual”¹¹. Para a Matemática, é uma perspectiva pior que para a Economia, pois reduz a Matemática ao papel secundário de abonadora do rigor e da precisão, espalhando as contribuições efetivas dessa ciência para a

Economia e vice-versa. Nem mesmo as contribuições do próprio Debreu se reduzem a apenas esse aspecto.

Por exemplo, embora Von Neumann destaque a clareza e a precisão, essas não são para ele as características mais importantes da Matemática:

[...] é perfeitamente claro que a matemática fornece algo que é bastante importante, a saber, estabelece certos padrões de objetividade, certos padrões de verdade; é muito importante que pareça fornecer um meio para estabelecer esses padrões independentemente de tudo o mais, independentemente de questões emocionais e morais. É muito importante alcançar essa percepção: critérios objetivos da verdade são possíveis, que tal objetivo não é autocontraditório, não é, em certo sentido, não humano. [...] Houve flutuações muito sérias na opinião dos matemáticos sobre o que é rigor matemático. Para mencionar uma coisa menor: na minha própria experiência, que se estende por apenas trinta anos, flutuou tão consideravelmente que minha convicção pessoal e sincera quanto ao rigor matemático mudou pelo menos duas vezes. [...] uma das contribuições mais importantes da matemática para o nosso pensamento é que ela tem demonstrado uma enorme flexibilidade na formação de conceitos, um grau de flexibilidade muito difícil chegar através de um modo não matemático.¹²

Tanto para von Neumann quanto para Morgenstern, a mera matematização de uma teoria econômica prévia pode não ter nenhum valor: “Frequentemente, é mais fácil matematizar uma teoria falsa que enfrentar a realidade.”¹³ A matematização é simultânea ao processo de elaboração da teoria. No *“Game Theory”*, von Neumann observa: “achamos que seja necessário recorrer a técnicas de matemática que não tenham sido utilizadas até agora em economia matemática, e é bem possível que um estudo mais aprofundado possa resultar, no futuro, na criação de novas disciplinas matemáticas.”¹⁴

Ainda devemos destacar que o pensamento matemático como fonte de intuições frutíferas não se relaciona com a dificuldade. O artigo *“The Distribution of a Product from Several Sources to Numerous Localities”* foi publicado pelo matemático norte-americano Frank

Lauren Hitchcock¹⁵ (1875-1957) em 1941. O trabalho foi redigido de maneira extremamente clara e não utiliza e nem adianta nenhum resultado matemático esotérico, ao contrário, a matemática empregada não ultrapassa a aritmética elementar. Sem dúvida, não foi o requinte que justificou a reputação do artigo. A fama justifica-se primeiramente por ser o primeiro trabalho a formular matematicamente um problema na área de transportes, tema muito sensível durante períodos de guerra, e em segundo lugar por propor um método de solução simples, viável, extremamente criativo que antecipou o chamado “Método Simplex”. O artigo de Hitchcock é um exemplo perfeito para a afirmação do Karl Popper acima.

O conteúdo deste livro não cobre toda a Economia Matemática. Os desenvolvimentos relativos à Dinâmica Econômica não estão aqui por exigirem um outro livro para seu tratamento adequado. Não tratamos também da existência do equilíbrio walrasiano por duas razões: por ter que desenvolver um instrumento matemático mais complexo para tratar de pontos fixos e por temor de desfocar a atenção do Teorema de Kuhn-Tucker. Ainda que a Teoria do Equilíbrio Geral seja uma teoria estática e comumente tratada junto com essa, acreditamos que é mais adequado tratá-la na Dinâmica.

Também não tratamos o importante tema das matrizes não-negativas, do Teorema de Frobenius e do Teorema de Hawkins-Simon, fundamentais para a compreensão adequada de Sraffa e das análises matemáticas do “Capital” de Marx, como proposto por Michio Morishima. Outra vez, esses temas precisam ser tratados num livro à parte, voltado exclusivamente para tais assuntos.

Schumpeter testemunhou uma metamorfose teórica:

[...] de uma coisa podemos estar certos. A teoria econômica de nosso tempo, como a das épocas futuras, jamais poderá ser outra vez tão fascinante para o grande público como o foi no tempo em que era compreensível para qualquer pessoa educada e parecia estabelecer diretamente “leis eternas” e regras práticas. Qualquer pessoa pode entender

Adam Smith. Apenas especialistas podem entender cálculo matricial e equações funcionais.¹⁶

Se essa tendência se sustentará, apenas o futuro poderá nos informar. O objetivo deste livro é auxiliar o leitor a se integrar no fluxo histórico.

-
- ¹ POPPER, Karl. **Autobiografia Intelectual**. São Paulo: Editora Cultrix, Editora da Universidade de São Paulo, 1977, p. 3.
- ² VON NEUMANN, J. The Role of Mathematics in the Sciences and in Society. In TAUB, A. H. John von Neumann, **Collected Works**, v. VI. NY: Pergamon Press, 1963, p. 482
- ³ “Na minha opinião, o artigo de von Neumann [*A Model of General Economic Equilibrium*’. **Rev. Econ. Stud.**, v. 13, nº 33 (1945-1946)] é o mais importante artigo em Economia Matemática.”. (WEINTRAUB, R. “On the Existence of a Competitive Equilibrium: 1930-1954”, **Journal of Economic Literature**, v. 21, nº 1 (1983), p. 13).
- ⁴ DEBREU, Gerard. “Theoretic Models: Mathematical Form and Economic Content”. **Econometrica**, v. 54, nº 6 (nov./1986), p. 1259.
- ⁵ NEUMANN, J. “Zur Theorie der Gesellschaftsspiele”. **Mathematische Annalen**, v. 100, nº 1 (1928), pp. 295–320. Esse artigo foi traduzido para o inglês por Sonya Bargmann com o título “*On the Theory of Games of Strategy*” e abre o volume IV do “*Contributions to The Theory of Games*”, nº 40 dos “*Annals of Mathematic Studies*” de 1959.
- ⁶ HANCOCK, H. **Theory of Maxima and Minima**. EUA: Ginn Company, 1917, pp. 115-116.
- ⁷ DEBREU, G. “Definite and Semidefinite Quadratic Forms”. **Econometrica**, v. 20, nº 2 (1952), pp. 295-300.
- ⁸ ARROW, K.; HURWICZ, L.; UZAWA, H. “Constraint Qualifications in Maximization Problems”. **Naval Research Logistics**, v. 8, nº 2 (1961).
- ⁹ ARROW, K.; ENTHOVEN, A. C. “Quasi-Concave Programming”. **Econometrica**, v. 29, nº 4 (1961), pp. 779-800.
- ¹⁰ DEBREU, G. “Theoretic Models: Mathematical Forms and Economic Content”. **Econometrica**, v. 54, nº 6 (1986), pp. 1259-1270.
- ¹¹ LEONTIEF, W. “The State of Economic Science”. **The Review of Economics and Statistics**, v. 40, nº 2 (1958), p. 104.
- ¹² NEUMANN, J. “The Role of Mathematics in the Sciences and in Society”. In: NEUMANN, J. **Collected Works**, v. VI. New York: The Macmillan Company, 1963.
- ¹³ MORGENSTERN, Oskar. “Thirteen Critical Points in Contemporary Economic Theory: An Interpretation”. **Journal of Economic Literature**, v. 10, nº 4 (dez./1972), pp. 1163-1189.

¹⁴ NEUMANN, J.; MORGENSTERN, O. **Theory of games and Economic Behavior**. 7ª ed. Princeton: Princeton University Press, 1953.

¹⁵ HITCHCOCK, F. L. "The Distribution of a Product from Several Sources to Numerous Localities". **Journal of Mathematics and Physics**, nº 20 (1941), pp. 224-230.

¹⁶ SCHUMPETER, Joseph. **História da Análise Econômica**. Rio de Janeiro: Editora Fundo de Cultura, 1964, p. 1110.

Lição UM

A Programação Linear

Sumário

Lição 1 - A Programação Linear

1. Introdução
2. O simpósio número zero da Programação Linear
3. O teorema fundamental
4. Problema 1: Análise de Atividade
 - 4.1 O problema dual (definição de “preços”)
 - 4.2 O problema primal (a Análise de Atividade)
 - 4.3 Tjalling Charles Koopmans
5. Problema 2: o Problema da Dieta
6. Problema 3: o Problema de Transporte
 - 6.1 A solução de Hitchcock
7. O problema C1 de Leonid Kantorovich
 - 7.1 Leonid Vitalevich Kantorovich
8. A programação vetorial: caso linear
 - 8.1 Definição de mínimos e máximos vetoriais
10. Comentários históricos sobre otimização vetorial

Lição 1

A Programação Linear

1. Introdução

O objetivo desta lição é apresentar a Programação Linear. O resultado básico é intuitivamente óbvio: se existem pontos de mínimo ou máximo para funções lineares definidas em polígonos ou poliedros, então existem pontos de mínimo ou máximo nos vértices dos polígonos ou poliedros.

As restrições que definem o problema são determinadas por semiespaços definidos por hiperplanos. Todo hiperplano fraciona o espaço em dois semiespaços: $a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n \leq b$ e $a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n \geq b$. A restrição Ω é determinada pela interseção de semiespaços. Caso existam mínimos ou máximos da função objetivo $f(x_1, x_2, \dots, x_n) = c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n$ quando restrita ao Ω , tanto os mínimos quanto os máximos estão sobre isoquantas $f(x_1, x_2, \dots, x_n) = c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n = c$ que dividem o espaço de maneira que a restrição Ω está inteiramente num dos semiespaços definido pelas isoquanta que contém esses mínimos ou máximos. Essa isoquanta contém uma aresta ou pelo menos um vértice do poliedro. Se o poliedro for limitado, então sempre existem mínimos e máximos nos vértices.

Elucidamos o resultado acima através de quatro exemplos fundamentais da Programação Linear: o problema de Análise de Atividades, o Problema da Dieta, o problema dos transportes e o

problema de encontrar as estratégias ótimas em jogos de soma-zero. Além desses problemas, apresentamos resumidamente o chamado Problema C1, do matemático russo Leonid Kantorovich (1912-1986).

2. O simpósio número zero da Programação Linear

O primeiro congresso dedicado exclusivamente à Programação Linear aconteceu em 1949, em Chicago. Foi chamado de “Simpósio Número Zero”. O décimo nono, por exemplo, ocorreu no Rio de Janeiro, na Universidade Federal do Rio de Janeiro (UFRJ), em 2006¹.

O “Simpósio Número Zero” resultou de uma reunião, em 1948, entre Tjalling Koopmans (1910-1985), George Dantzig (1914-2005), Harold Kuhn (1925-2014), Albert Tucker (1905-1995), Oskar Morgenstern (1902-1977) e Wassily Leontief (1906-1999). Na época, Koopmans era o diretor de pesquisa da Comissão Cowles para Pesquisa em Economia e ficou encarregado de organizar o encontro, ou seja, obter o dinheiro necessário que veio da RAND Corporation².

A introdução aos “Proceedings” do simpósio, que foi escrita pelo Koopmans, esclarece:

As contribuições para este livro estão voltadas, direta ou indiretamente, a vários aspectos do problema fundamental da economia normativa: a melhor alocação de meios limitados em relação a finalidades desejadas. O lugar central deste problema no pensamento econômico esclarece o que de outra maneira pode parecer um fato surpreendente: que os estudos aqui reunidos exibem afinidades e conexões entre linhas de pensamento e reservas de experiência desenvolvidas aparentemente de maneira independente por vários grupos de economistas, matemáticos e administradores.³

O Simpósio Zero foi revolucionário pela abordagem multidisciplinar e pelas consequências para a Teoria Econômica, ainda que eventualmente esse não fosse o objetivo dos participantes⁴. Pela primeira vez, profissionais de áreas diferentes se

reuniam e combinavam suas especialidades na solução de problemas econômicos.

Os trabalhos pioneiros em Programação Linear são os artigos *“Mathematical Methods of Production Planning and Organization”*⁵ (1939), de Leonid Kantorovich (1912-1986), e *“The Distribution of a Product from Several Sources to Numerous Localities”*⁶, publicado em 1941 por Frank Lauren Hitchcock (1875-1957). Tjalling C. Koopmans⁷ (1910-1985) formulou o mesmo problema que Hitchcock em 1947. Tjalling Koopmans⁸ e Robert Dorfmann⁹ fazem referência à primazia de Hitchcock. Ainda, segundo Koopmans, Maurice Allais (1911-2010) também formulou o mesmo problema.

O Problema da Dieta foi formulado por George Stigler¹⁰ (1911-1991) em 1945, que não o solucionou utilizando os métodos da Programação Linear.

A Análise de Atividades resultou das investigações de George Dantzig¹¹ (1914-2005) para a Força Aérea Norte-Americana na década de 1940. Inicialmente, a Análise de Atividades se relacionava com questões de abastecimento, transporte e coordenação de operações em planejamento militar e depois da 2ª Guerra Mundial teve aplicações nas atividades civis. George Dantzig reconheceu o padrão comum em vários desses problemas, criando a disciplina que chamamos Programação Linear. A expressão “Programação Linear” deve-se a Tjalling Koopmans e se origina do jargão militar.

Em 1947, Dantzig desenvolveu o Método Simplex para resolução dos problemas de Programação Linear.

Von Neumann foi o primeiro a perceber a relação entre a Programação Linear e a Teorias dos Jogos¹² e expressou suas ideias para Dantzig no primeiro encontro dos dois, em Princeton, no ano de 1947. A relação entre jogos de soma-zero e a Programação Linear é

explorada na Parte Três dos “*Proceedings*” do “Simpósio Zero”. No artigo “*A Proof of the Equivalence of the Programming Problem and the Game Theory*”, capítulo XX dos “*Proceedings*”, Dantzig enfatiza que a intuição original foi de von Neumann¹³.

Em 1951, foi publicado o livro “*Activity Analysis of Production and Allocation*”, editado por Tjalling Koopmans, que contém parte das conferências do congresso número zero.

Na introdução do livro, Koopmans atribui as origens da Programação Linear primeiramente às contribuições feitas à teoria do equilíbrio por alguns frequentadores, principalmente Abraham Wald (1902-1950) e von Neumann (1903-1957), nos famosos seminários organizados por Karl Menger (1902-1985) em Viena, na década de 1930. Em segundo lugar, ainda nos seminários, Koopmans lembra as discussões sobre a Economia do bem-estar conduzidas por Ludwig von Mises (1881-1973) e Oscar Lange (1904-1965). Em terceiro lugar, temos a análise de insumo-produto feita por Wassily Leontief (1906-1999) na década de 1930. Finalmente, em quarto lugar, Koopmans coloca os problemas logísticos originados nas atividades bélicas. No entanto, Koopmans sublinha que a Programação Linear se insere dentro do contexto maior da teoria econômica, pois trata “da melhor alocação de recursos limitados em relação a fins desejados”.

Os matemáticos americanos desconheciam desenvolvimentos similares na União Soviética por parte da equipe de Leonid Kantorovich.

Em 1975, Koopmans dividiu com Leonid Kantorovich o prêmio Nobel pelas contribuições dadas pelos dois à Economia através da Programação Linear. Segundo se diz, houve certa surpresa e indignação nos meios acadêmicos norte-americanos por não ter sido Dantzig premiado conjuntamente. No mesmo ano, Dantzig recebeu a

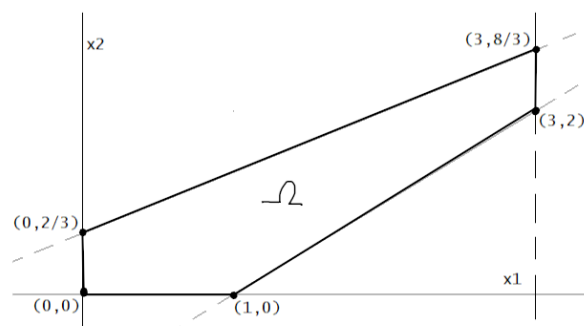
“National Medal of Science” e o “John von Neumann Theory Prize” como demonstração de desagravo pela decisão dos suecos¹⁴: “Infelizmente, para seus admiradores na comunidade da Programação Linear, o prêmio Nobel de 1975 foi para L. V. Kantorovich e T. C. Koopmans. Alguns viram motivação geopolítica nesta seleção.”¹⁵

3. O teorema fundamental

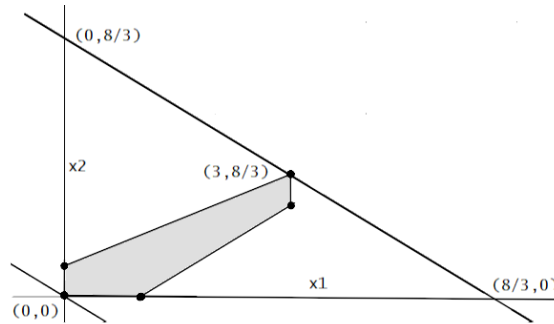
Exemplo 1: Consideremos o polígono Ω definido pelas inequações abaixo:

$$\begin{aligned} x_1 - x_2 &\leq 1 \\ -2 \cdot x_1 + 3 \cdot x_2 &\leq 2 \\ x_1 &\leq 3 \\ x_1 &\geq 0, x_2 \geq 0. \end{aligned}$$

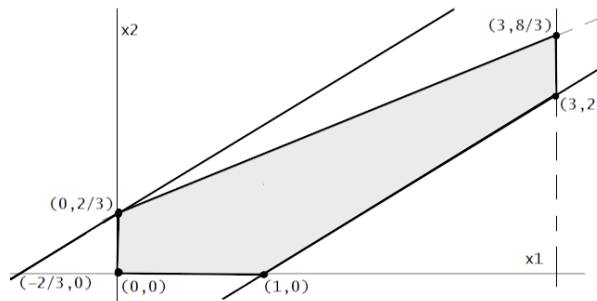
O polígono e seus vértices estão ilustrados abaixo:



Digamos que a função objetivo seja $f(x_1, x_2) = x_1 + x_2$, então a região Ω está à esquerda da isoquanta $f(x_1, x_2) = x_1 + x_2 = \frac{17}{3}$, que é a isoquanta de maior nível que intercepta a região Ω no vértice $(3, \frac{3}{8})$; a isoquanta de nível menor é $f(x_1, x_2) = x_1 + x_2 = 0$ que intercepta a região Ω no ponto $(0,0)$; a região está à direita da isoquanta. Veja abaixo:



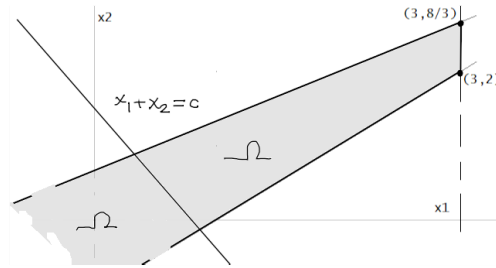
Caso a função objetivo seja $f(x_1, x_2) = x_1 - x_2$, a região Ω está entre as isoquantas $x_1 - x_2 = -\frac{2}{3}$ e $x_1 - x_2 = 1$. O ponto de mínimo está no vértice $(0, \frac{2}{3})$ e a aresta entre os vértices $(0, \frac{2}{3})$ e $(3, 2)$ é constituída de máximos. Veja o desenho abaixo:



Pode ocorrer também que mínimos ou máximos não existam. Por exemplo, a região Ω determinada por

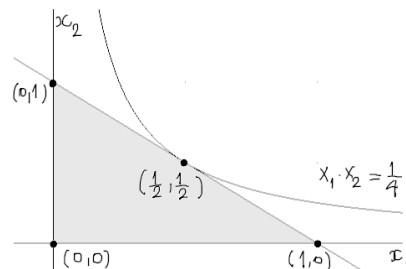
$$\begin{aligned} x_1 - x_2 &\leq 1 \\ -2 \cdot x_1 + 3 \cdot x_2 &\leq 2 \\ x_1 &\leq 3, \end{aligned}$$

tem apenas dois vértices: $(3, \frac{8}{3})$ e $(3, 2)$. Se a função objetivo é $f(x_1, x_2) = x_1 + x_2$, então o máximo está no vértice $(3, \frac{8}{3})$ e a região está à esquerda da isoquanta $f(x_1, x_2) = x_1 + x_2 = \frac{17}{3}$, como vimos acima. Mas, observe que não existem mínimos para essa função. Não existe nenhuma isoquanta $f(x_1, x_2) = x_1 + x_2 = c$ tal que a região fique totalmente à direita da isoquanta. Todas cortam a região.



Desses exemplos, devemos concluir que, caso existam mínimos ou máximos para a função objetivo, existe pelo menos um ponto de mínimo ou de máximo num vértice do poliedro. Esse é o resultado que empregamos nos exemplos a seguir.

Exemplo 2: Não é verdade que sempre existem pontos de mínimo ou de máximo nos vértices dos polígonos. Se o problema é maximizar a função $f(x_1, x_2) = x_1 \cdot x_2$, quando restrita ao triângulo $x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0$, então o ponto de máximo está em $(\frac{1}{2}, \frac{1}{2})$ e não nos vértices $(0,0)$, $(1,0)$ ou $(0,1)$. Veja a figura:



4. Problema 1: Análise de Atividade

Exemplo 3: Suponhamos que temos dois produtos produzidos através de dois processos e pela utilização de três insumos. O primeiro processo emprega 1 unidade do insumo 1, 1 unidade do insumo 2 e 2 unidades do insumo 3 para produzir uma unidade do produto 1. Portanto, para produzir x_1 unidades do produto 1, necessitamos x_1 unidades do insumo 1, x_1 unidades do insumo 2 e $2 \cdot x_1$ unidades do insumo 3. O segundo processo emprega 3 unidades do

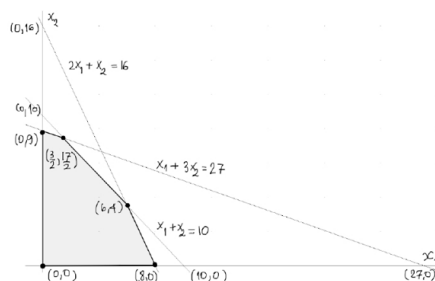
insumo 1, 1 unidade do insumo 2 e 1 unidade do insumo 3 para produzir 1 unidade do segundo produto. Portanto, para produzir x_2 unidades desse segundo produto, necessitamos $3 \cdot x_2$, x_2 e x_2 unidades dos insumos 1, 2 e 3, respectivamente. As quantidades disponíveis destes insumos são, 27, 10 e 16 unidades, respectivamente. Observe a tabela abaixo:

	Produto 1	Produto 2	disponibilidade
Unidades de Insumo 1/unidade de produto	1	3	27
Unidades de Insumo 2/unidade de produto	1	1	10
Unidades de Insumo 3/unidade de produto	2	1	16

Portanto, a estrutura produtiva e as restrições de insumos são expressas pelas inequações:

$$\begin{aligned} x_1 + 3 \cdot x_2 &\leq 27 \\ x_1 + x_2 &\leq 10 \\ 2 \cdot x_1 + x_2 &\leq 16 \\ x_1 \geq 0, x_2 &\geq 0. \end{aligned}$$

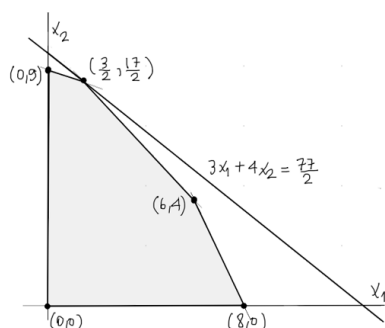
Essas inequações determinam a região do plano $\Omega \subseteq \mathbb{R}^2$, que está em cinza no gráfico abaixo. Essa região Ω é chamada região viável, pois contém as combinações possíveis de insumos para essa estrutura tecnológica. Os vértices do polígono Ω são $(0,0)$, $(0,9)$, $(3/2, 17/2)$, $(6,4)$ e $(8,0)$.



Suponhamos que cada unidade do produto 1 é vendida por R\$ 3,00 e cada unidade do produto 2, por R\$ 4,00. Então, a receita advinda da venda de x_1 e x_2 unidades desses produtos é $R = 3 \cdot x_1 + 4 \cdot x_2$. Nosso objetivo é descobrir no polígono Ω acima qual a combinação de produtos (x_1, x_2) que maximiza a receita R . Em outras palavras, resolver o problema:

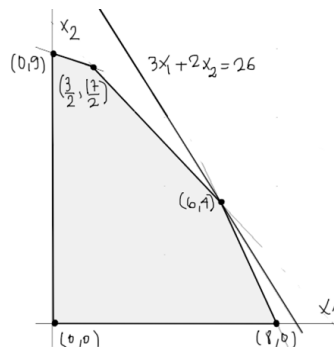
$$\begin{array}{ll} \max & 3 \cdot x_1 + 4 \cdot x_2 \\ \text{sa} & x_1 + 3 \cdot x_2 \leq 27 \\ & x_1 + x_2 \leq 10 \\ & 2 \cdot x_1 + x_2 \leq 16 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

em que *sa* significa “sujeito às restrições” que estão à direita. Consideremos as isoquantas da função objetivo que são as retas $3 \cdot x_1 + 4 \cdot x_2 = R$. Nosso objetivo é obter a isoquanta com o R maior possível que intercepta o polígono Ω . Essa reta está desenhada abaixo, $R = \frac{77}{2}$, e intercepta a região Ω no vértice $(\frac{3}{2}, \frac{17}{2})$. Isso ilustra nossa observação anterior: sempre temos pelo menos uma solução num vértice.



Portanto, se queremos obter uma solução, basta que calculemos a função objetivo em cada um dos vértices e selecionemos o maior valor. Observe que o insumo 3 não é integralmente utilizado. A solução ótima $(\frac{3}{2}, \frac{17}{2})$ é tal que $2 \cdot (\frac{3}{2}) + (\frac{17}{2}) = \frac{23}{2} < 16$.

Por exemplo, se $f(x_1, x_2) = 3 \cdot x_1 + 4 \cdot x_2$, então $f(0,0) = 0$, $f(0,9) = 18$, $f(\frac{3}{2}, \frac{17}{2}) = \frac{43}{2}$, $f(6,4) = 26$ e $f(8,0) = 24$; disso concluímos que o máximo está em $f(6,4) = 26$.

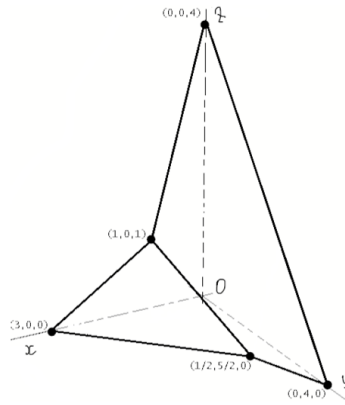


Chamamos esse problema que estamos tratando de **problema primal**.

Exemplo 4: Vejamos outro problema, o chamado **problema dual**. Suponhamos que queremos atribuir preços aos insumos empregados no problema anterior. Digamos que para cada unidade dos insumos 1, 2 e 3, atribuímos os preços y_1 , y_2 e y_3 , respectivamente. Observemos que, com esses preços, o custo para produzir uma unidade do produto 1 é $y_1 + y_2 + 2 \cdot y_3$, pois necessitamos 1 unidade do insumo 1, 1 unidade do insumo 2 e 2 unidades do insumo 3. Da mesma maneira, o custo para produzir 1 unidade do produto 2 é $3 \cdot y_1 + y_2 + y_3$. Como as unidades desses produtos são vendidas, por exemplo, por R\$ 3,00 e por R\$ 4,00, respectivamente, e as receitas devem cobrir os custos, então $y_1 + y_2 + 2 \cdot y_3 \geq 3$ e $3 \cdot y_1 + y_2 + y_3 \geq 4$, no mínimo. Como o total de insumos disponíveis são 27 unidades, 10 unidades e 16 unidades, então o custo total é $C = 27 \cdot y_1 + 10 \cdot y_2 + 16 \cdot y_3$. Obviamente, queremos minimizar C . Portanto, temos um novo problema:

$$\begin{array}{ll} \min & 27 \cdot y_1 + 10 \cdot y_2 + 16 \cdot y_3 \\ \text{sa} & y_1 + y_2 + 2 \cdot y_3 \geq 3 \\ & 3 \cdot y_1 + y_2 + y_3 \geq 4 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, \end{array}$$

O chamado problema dual. Esse problema é levemente mais complicado, porque as restrições determinam uma região que é exterior – a região viável – ao poliedro desenhado abaixo, e cujos vértices estão indicados:



Novamente percebemos que deve existir uma solução num vértice, ou seja, a isoquanta de nível mais baixo da função objetivo $f(y_1, y_2, y_3) = 27 \cdot y_1 + 10 \cdot y_2 + 16 \cdot y_3 = C$ intercepta a região viável num dos vértices. Como, $f(3, 0, 0) = 81$, $f(1, 0, 1) = 43$, $f(0, 0, 4) = 64$, $f(0, 4, 0) = 40$ e $f(\frac{1}{2}, \frac{5}{2}, 0) = \frac{77}{2}$, constatamos que a solução está no vértice $(\frac{1}{2}, \frac{5}{2}, 0)$.

Dois fatos que não nos surpreendem: primeiramente, que o valor ótimo no problema primal e no problema dual é o mesmo, $\frac{77}{2}$, ou seja, que a receita máxima é igual ao custo mínimo; em segundo lugar, observe que o preço do insumo 3 é zero. Isso ocorre porque esse insumo está disponível em maior quantidade do que aquela que foi utilizada.

Os dois problemas acima, do ponto de vista formal, se expressam como:

4.1 O problema dual (definição de “preços”)

Associado ao problema acima é:

$$\begin{array}{ll} \min & b_1 \cdot y_1 + \dots + b_m \cdot y_m \\ \text{sa} & a_{11} \cdot y_1 + \dots + a_{m1} \cdot y_m \geq c_1; \\ & \vdots \quad \quad \quad \vdots \\ & a_{1n} \cdot y_1 + \dots + a_{mn} \cdot y_m \geq c_n; \\ & y_1 \geq 0, \dots, y_m \geq 0; \end{array}$$

Observe que em termos matriciais:

$$\begin{array}{ll} \max & b' \cdot y \\ \text{sa} & A' \cdot y \geq c; \\ & y \geq 0; \end{array}$$

em que A' é matriz transposta da matriz A . Do ponto de vista matemático, esses dois problemas foram estudados detalhadamente na década de 1950 pelo grupo de matemáticos ligados a von Neumann (1903-1957): Harold Kuhn (1925-2014), David Gale (1921-2008), George Dantzig (1914-2005) e Albert Tucker (1905-1995).

Num dos mais longos artigos do “*Activity Analysis*”¹⁶, Koopmans expressa o que entende por “Análise de Atividade”. Para Koopmans, o conceito de função de produção é limitado, pois a técnica usada na produção também envolve decisões gerenciais com o objetivo de maximizar o produto, que implicam escolhas de acordo com as diferentes disponibilidades de insumos. Além disso, na formulação das “atividades” existe a possibilidade da produção conjunta. Por isso, Koopmans prefere falar de “funções de transformações” no lugar de “função de produção”. Ademais, afirma que o modelo pode ser formulado de uma maneira dinâmica, e dirige o leitor para um artigo de Dantzig no capítulo II do livro.¹⁷

De fato, Koopmans tem em mente a possibilidade de expressar toda a Economia através de formidáveis problemas primais e duais, não exatamente como no nosso exemplo, mas de uma maneira um pouco mais complexa envolvendo programação vetorial, como veremos mais adiante. Koopmans deixa isso claro no prefácio do livro:

As contribuições para este livro estão voltadas, direta ou indiretamente, aos vários aspectos de um problema fundamental da economia normativa: a melhor alocação de meios limitados para finalidades desejadas. O lugar central deste problema central no pensamento econômico expõe o que, caso contrário, pode parecer surpreendente: que os estudos aqui reunidos exibem conexões e afinidades entre linhas de pensamento e experiências desenvolvidas em aparente independência por diversos grupos de economistas, matemáticos e administradores.¹⁸

Como precursores dessa série de trabalhos, Koopmans enumera primeiramente os esforços feitos na década de 1930, em Viena, para formular de maneira coerente o sistema de Equilíbrio Geral Walrasiano, que culminaram com uma prova matemática da existência de equilíbrio feita pelo matemático Abraham Wald¹⁹ (1902-1950). Koopmans destaca que nessa etapa surge o conceito de bens escassos e de bens livres que aparecem no exemplo acima. Esses conceitos são necessários para não obter preços negativos como soluções do sistema. Depois, Koopmans destaca o artigo de von Neumann²⁰, de 1937, que incorporou num modelo dinâmico único as atividades que permanecem e aquelas que são interrompidas. Nesse ponto, Koopmans lembra que, tanto nas formulações de Wald quanto nas de von Neumann, em mercados competitivos, os lucros se anulam. Um resultado que obtemos no nosso exemplo. Novamente, essa formulação se vincula à necessidade de obter preços maiores ou iguais a zero como soluções.

Como segunda fonte dos trabalhos expostos no livro, Koopmans destaca a função objetivo, introduzida por Dantzig, e os conceitos de ótimo, introduzidos por Pareto, que vinculam a Teoria do Bem-Estar. Nesse momento, Koopmans lembra as discussões sobre o “Cálculo Socialista”, ou seja, sobre as possibilidades de os preços-sombra servirem como índices de informação, alocação e distribuição de bens e de como os preços são usados numa economia capitalista.

A terceira fonte citada por Koopmans é Leontief, com a formulação das equações que expressam as atividades interindustriais. Koopmans está se referindo à possibilidade de obter empiricamente os dados dessas relações, como Leontief havia feito para economia americana²¹. Essa possibilidade é essencial para que os modelos reunidos no livro tenham emprego efetivo.

Finalmente, Koopmans faz referência aos trabalhos conduzidos pelo grupo – que incluía George Dantzig – sob a direção de um militar, Marshall Wood, em relação ao planejamento das atividades militares:

O trabalho mais abrangente desta categoria foram o desenvolvimento de modelos de programação por George Dantzig [...] vários modelos deste grupo estão neste livro. [...] Os modelos deste grupo foram desenvolvidos para tratar como os aspectos dinâmicos do planejamento das atividades interdependentes de grandes organizações e com a melhor combinação de atividades com a finalidade de obter um objetivo estabelecido.²²

4.2 O problema primal (a Análise de Atividade)

$$\begin{array}{ll} \max & c_1 \cdot x_1 + \dots + c_n \cdot x_n \\ \text{sa} & a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n \leq b_1; \\ & \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ & a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n \leq b_m; \\ & x_1 \geq 0, \dots, x_n \geq 0; \end{array}$$

na forma matricial representaremos o problema acima como:

$$\begin{array}{ll} \max & c' \cdot x \\ \text{sa} & A \cdot x \leq b; \\ & x \geq 0; \end{array}$$

$$\text{em que } c' = (c_1 \quad \dots \quad c_n), A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \text{ e } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix};$$

4.3 Tjalling Charles Koopmans

Tjalling Charles Koopmans (1910-1985) terminou seu Ph.D. em Leiden (Holanda), em 1935, sob a supervisão de Hans Kramers (físico teórico) e Jan Tinbergen (1903-1994) (estatístico). A tese de Koopmans, intitulada “*Linear Regression Analysis of Economic Time Series*”, foi publicada em 1936 e foi parcialmente concluída em Oslo, enquanto estudava com o economista Ragnar Anton Kittil Frisch²³ (1895-1973). Ainda na Holanda, Koopmans foi atraído pelo marxismo, o que o aproximou de Tinbergen.

Tinbergen e Frisch dividiram o primeiro Prêmio Nobel de Economia, em 1969. Tinbergen era físico de formação, mas destacou-se igualmente como econometrista. Frisch foi o criador da expressão “Econometria” num artigo publicado em 1926.

De 1938 a 1940, Koopmans substituiu Tinbergen, em Genebra, no projeto de pesquisa em ciclos econômicos para a Liga das Nações. Em 1939, Koopmans publicou o livro *“Tanker Freight Rates and Tankship Building: An Analysis of Cyclical Fluctuations”*, onde analisa a demanda mundial por fretamento de navios petroleiros. Esse trabalho motivou outro, feito para os Aliados em 1942, intitulado: *“Exchange Ratios between Cargoes on Various Routes (Non-Refrigerated Dry Cargoes)”*, sobre navegação mercante durante a 2ª Guerra e contribuiu para a sua premiação com o Nobel²⁴:

O melhor exemplo do uso da otimização em operações navais durante a Segunda Guerra Mundial foi o trabalho de Tjalling J. Koopmans, vencedor do Prêmio Nobel de Ciências Econômicas de 1975. Enquanto servia como estatístico no Combined Shipping Adjustment Board em Washington, ele foi convidado a estudar a rota ideal de navios aliados vazios. Para resolver esse problema, ele aplicou uma forma inicial de programação linear. Sua solução enfatizou a importância dos preços sombra e propôs uma recomendação numérica sobre o que transportar e para onde.²⁵

Na monografia n. 13 da Fundação Cowles, que constitui os *Proceedings* da Conferência Número Zero, um dos trabalhos apresentados por Koopmans é o artigo *“A Model de Transportation”*.

Em 1944, Koopmans foi para os EUA fugindo – como tantos – do nazismo e ingressou na Comissão Cowles a convite de Jacob Marschak (1898-1977). Em 1948, foi nomeado diretor de pesquisa da Fundação Cowles; vice-presidente da Econometric Society, em 1949, e presidente em 1950. Enquanto a Fundação Cowles esteve alocada na Universidade de Chicago, Koopmans também estava, mas, em 1955, mudou-se para a Universidade de Yale acompanhando a Comissão Cowles.

Em 1947, Koopmans publicou o artigo *"Measurement Without Theory"*, que é uma crítica à metodologia adotada pelos institucionalistas Wesley C. Mitchell (1875-1948) e Arthur F. Burns (1904-1987) na análise de ciclos econômicos. O artigo causou uma acirrada polêmica entre institucionalistas e neoclássicos. Em 1957, Koopmans publicou o livro *"Three Essays on State of Economic Science"*, que novamente causou reações.

Quatro trabalhos de Koopmans são considerados contribuições importantes para a Econometria: o artigo *"Identification Problems in Economic Model Construction"*, publicado na *Econometrica* em 1949 e que também aparece na monografia n. 14 de 1953 da Fundação Cowles; *"When is an Equation System Complete for Statistical Purposes?"* e *"Measuring the Equation System of Dynamic Economics"*, que se encontram na monografia n. 10 (1950) da Fundação Cowles; e *"The Estimation of Simultaneous Linear Economic Relationships"*, que se destaca na monografia n. 14 (1953)²⁶. Carl Christ considera Koopmans um dos responsáveis pela chamada "Revolução Econométrica" ocorrida no seio da Fundação Cowles²⁷.

Koopmans compartilhou o Prêmio Nobel com Kantorovich por sua pesquisa sobre a alocação ótima de recursos em 1975.

Uma informação que não é usual que os economistas saibam: a carreira de Koopmans começou, em 1934, com um resultado célebre em Química Quântica conhecido como "Teorema de Koopmans"²⁸.

5. Problema 2: o Problema da Dieta

O Problema da Dieta foi formulado por George Stigler (1911-1991) no artigo *"The Cost of Subsistence"*²⁹ de 1945. Stigler recebeu o Nobel em 1982. Conforme nota da página 9 do livro *"Linear Programming and Economic Analysis"* (1958), de Robert Dorfman, Paul Samuelson e

Robert Solow, o Problema da Dieta já havia sido tratado, em 1941, por um eminente bioestatístico chamado Jerome Cornfield (1912-1979) num trabalho não publicado³⁰. A solução, via Programação Linear, foi obtida por George Dantzig e um colaborador em 1947, mas também não foi publicada:

A primeira aplicação em grande escala do método simplex foi a determinação de uma dieta adequada pelo menor custo. No outono de 1947, J. Laderman do Projeto de Tabelas Matemáticas da Agência Nacional de Padrões (National Bureau of Standards) supervisionou a solução de nove equações em setenta e sete variáveis não-negativas. Utilizando calculadoras de mesa, seriam necessários 120 homens por dia. [...] O problema resolvido foi aquele estudado anteriormente por G. Stigler que conjecturou uma solução apenas 24 centavos maior que o custo mínimo verdadeiro de \$39,69 para um ano.³¹

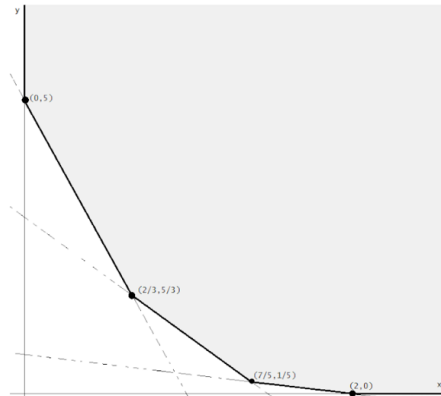
Exemplo 5: Digamos que um jardineiro tem à disposição duas marcas de adubo, digamos A e B. Ambas as marcas possuem os nutrientes I, II e III necessários para a conservação da fertilidade do solo. As informações sobre A e B estão na tabela abaixo:

Nutrientes	Adubo A	Adubo B	Quantidades necessárias mínimas de nutrientes para cada 1 m ² de superfície de solo
I	2 g/kg	1 g/kg	3 g/m ²
II	1 g/kg	3 g/kg	2 g/m ²
III	5 g/kg	1 g/kg	5 g/m ²
Custo/kg	R\$ 6,00	R\$ 4,00	

Cada 1 metro quadrado de solo necessita de quantidades mínimas de nutrientes que estão na quarta coluna da tabela acima; o custo por quilo de cada marca de adubo está na última linha. O problema do agricultor é combinar os dois adubos de maneira a satisfazer as necessidades do solo ao menor custo possível. Se x e y representam os números de quilos dos adubos A e B, respectivamente, então, matematicamente, o problema pode ser formulado como:

$$\begin{array}{ll}
 \min & 6 \cdot x + 4 \cdot y \\
 \text{sa} & 2 \cdot x + y \geq 3 \\
 & x + 3 \cdot y \geq 2 \\
 & 5 \cdot x + y \geq 5 \\
 & x \geq 0, y \geq 0.
 \end{array}$$

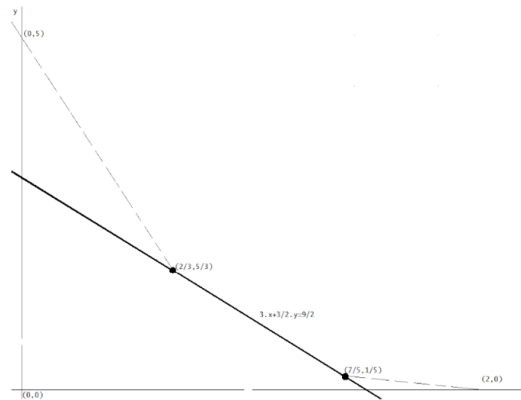
Abaixo em cinza está a região das combinações viáveis e os respectivos vértices.



Como existe pelo menos uma solução em algum vértice, comutando o valor do custo, $C(x, y) = 6 \cdot x + 4 \cdot y$, nesses vértices encontramos uma solução. No caso, a solução está em $(\frac{7}{5}, \frac{1}{5})$, em que o custo é o menor possível $C(\frac{7}{5}, \frac{1}{5}) = 9,2$. Observe o gráfico abaixo, em que a isoquanta $C(x, y) = 6 \cdot x + 4 \cdot y = \frac{46}{5}$ está desenhada:



Observe que se a função custo é $C(x, y) = 3 \cdot x + \frac{3}{2} \cdot y$, o menor custo está ao longo do segmento de reta que une os vértices $(\frac{2}{3}, \frac{5}{3})$ e $(\frac{7}{5}, \frac{1}{5})$, em que o custo mínimo é $\frac{9}{2}$. Observe o desenho abaixo:



Podemos descobrir em quais vértices estão as soluções analisando as declividades dos segmentos de reta que delimitam a região viável e a declividade da função-objetivo que estamos minimizando (ou maximizando). As declividades das retas $5 \cdot x + y = 5$, $2 \cdot x + y = 3$ e $x + 3 \cdot y = 2$ são -5 , -2 e $-\frac{1}{3}$, respectivamente. A reta de custo $C(x, y) = 6 \cdot x + 4 \cdot y$ tem declividade igual a $-\frac{3}{2}$, portanto, não é difícil ver que a solução estará entre a reta $x + 3 \cdot y = 2$ e $2 \cdot x + y = 3$ já que $-2 < -\frac{3}{2} < -\frac{1}{3}$, ou seja, no vértice $(\frac{7}{5}, \frac{1}{5})$. No caso em que o custo é $C(x, y) = 3 \cdot x + \frac{3}{2} \cdot y$, a declividade das retas de custo é -2 e, portanto, a reta de custo mínimo é a reta $2 \cdot x + y = 3$.

A formulação geral do Problema da Dieta é:

$$\begin{aligned} \min \quad & c_1 \cdot y_1 + \dots + c_m \cdot y_m \\ \text{sa} \quad & a_{11} \cdot y_1 + \dots + a_{1m} \cdot y_m \geq b_1; \\ & \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ & a_{n1} \cdot y_1 + \dots + a_{nm} \cdot y_m \geq b_n; \\ & y_1 \geq 0, \dots, y_m \geq 0; \end{aligned}$$

em que $c_1, \dots, c_m$ são os preços de cada unidade das rações $1, \dots, m$, respectivamente; $b_1, \dots, b_n$ são os números de unidades mínimas dos nutrientes $1, \dots, n$; a_{ij} é o número de unidades do nutriente i presente em uma unidade da ração j ; $y_1 \geq 0, \dots, y_m \geq 0$ são as unidades das rações $1, \dots, m$ a serem determinadas.

Observe que o Problema da Dieta é da forma

$$\begin{array}{ll}
 \min & b_1 \cdot y_1 + \dots + b_m \cdot y_m \\
 \text{sa} & a_{11} \cdot y_1 + \dots + a_{m1} \cdot y_m \geq c_1; \\
 & \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
 & a_{1n} \cdot y_1 + \dots + a_{mn} \cdot y_m \geq c_n; \\
 & y_1 \geq 0, \dots, y_m \geq 0,
 \end{array}$$

ou seja, de um problema dual.

Exercício 1

I. Determine os vértices do polígono

$$\begin{array}{l}
 3 \cdot x + 4 \cdot y \geq 12 \\
 3 \cdot x + 2 \cdot y \geq 8 \\
 2 \cdot x + 5 \cdot y \geq 10 \\
 x \geq 0, y \geq 0.
 \end{array}$$

Resposta: $(0, 4)$, $(4/3, 2)$, $(20/7, 6/7)$ e $(5, 0)$.

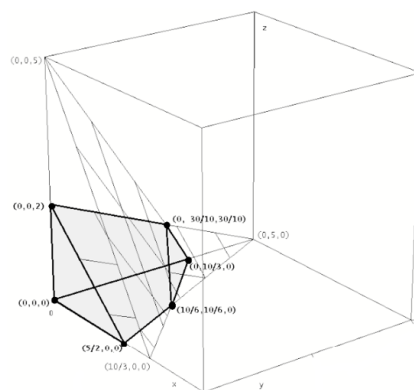
II. Resolva graficamente o problema:

$$\begin{array}{ll}
 \min & 10 \cdot x + 10 \cdot y \\
 & 3 \cdot x + 4 \cdot y \geq 12 \\
 \text{sa} & 3 \cdot x + 2 \cdot y \geq 8 \\
 & 2 \cdot x + 5 \cdot y \geq 10 \\
 & x \geq 0, y \geq 0.
 \end{array}$$

Observando que as declividades das retas que delimitam a região viável são $-\frac{3}{2} < -\frac{3}{4} < -\frac{2}{5}$, conclua que a solução ótima está no vértice $\left(\frac{4}{3}, 2\right)$.

Resposta: $f\left(\frac{4}{3}, 2\right) = 10 \cdot \frac{4}{3} + 10 \cdot 2 = \frac{100}{3}$.

III. Escreva o problema primal associado ao problema acima. Encontre os vértices do poliedro que determinam a região viável.



Resposta:

$$\begin{array}{ll}\max & 3 \cdot x_1 + 8 \cdot x_2 + 10 \cdot x_3 \\ \text{sa} & 3 \cdot x_1 + 3 \cdot x_2 + 2 \cdot x_3 \leq 10 \\ & 4 \cdot x_1 + 2 \cdot x_2 + 5 \cdot x_3 \leq 10 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0.\end{array}$$

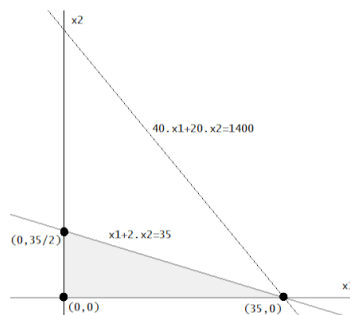
Observe que o valor máximo no problema primal é igual ao valor mínimo no problema dual.

Exercício 2

Como sabemos, as soluções dos problemas de maximização da utilidade, quando os bens são substitutos perfeitos, é uma “solução de canto”, ou seja, encontra-se num dos vértices do triângulo determinado pela restrição orçamentária. Na verdade, este é o problema mais simples entre os que estamos tratando. Resolva graficamente o problema:

$$\begin{array}{ll}\max & 40 \cdot x_1 + 20 \cdot x_2 \\ \text{sa} & x_1 + 2 \cdot x_2 \leq 35 \\ & x_1 \geq 0, x_2 \geq 0.\end{array}$$

Resposta:



Exercício 3

Considere o problema:

$$\max x + A \cdot y$$

sujeito à restrição

$$5 \cdot x + 2 \cdot y \leq 1,$$

$$2 \cdot x + 3 \cdot y \leq 1$$

$$x \geq 0, \quad y \geq 0.$$

$A > 0$. Sabendo que existe uma solução desse problema no ponto $(\frac{1}{11}, \frac{3}{11})$, determine se i) $A < \frac{2}{3}$; ii) $\frac{2}{3} \leq A < \frac{5}{2}$; iii) $A > \frac{5}{2}$. Justifique sua resposta.

Exercício 4

Considerando graficamente o problema:

$$\begin{array}{ll} \min & a \cdot x + y \\ & 3 \cdot x + y \geq 6 \\ & x + y \geq 4 \\ \text{sa} & x + 3 \cdot y \geq 6 \\ & x \geq 0, y \geq 0, \end{array}$$

sabendo que $2 < a < 3$, em qual vértice está a solução? Justifique a resposta através de um gráfico.

É óbvio que teremos de obter um processo mais eficaz para resolver esses problemas. Este método chama-se **Método Simplex** e foi uma criação do matemático George Dantzig.

Exercício 5

Três tipos de ração para cães, A, B e C, disponíveis no mercado, possuem os nutrientes I e II. As informações sobre A, B e C estão na tabela abaixo. Em relação à ração A, o gasto mensal de uma unidade de ração é R\$ 35,00; em relação à ração B, o gasto mensal de uma unidade é R\$ 24,00; e em relação à ração C, o gasto mensal de uma unidade é R\$ 28,00. Formule o Problema da Dieta e determine o problema dual. Não é preciso resolver o problema.

Nutrientes	Alimento A	Alimento B	Alimento C	Quantidades mensais
I	25	20	32	20 unidades
II	20	30	18	24 unidades
Custo	R\$ 35,00	R\$ 24,00	R\$ 28,00	

Exercício 6

Uma empresa química dispõe de 30 ton de nitrato e 80 ton de fosfato para produzir três tipos diferentes de fertilizantes. Na tabela abaixo estão os dados de uso e receita para cada um dos fertilizantes:

tipo de fertilizante	nitrato	fosfato	Receita
I	2 ton/1000 sacos	4 ton/1000 sacos	R\$ 30 /1000 sacos
II	3 ton/1000 sacos	4 ton/1000 sacos	R\$ 60 /1000 sacos
III	2 ton/1000 sacos	2 ton/1000 sacos	R\$ 50 /1000 sacos

Formule o problema de maximização da receita. Não é preciso resolver o problema.

Exercício 7

Uma firma produz dois bens nas quantidades x_1 ton e x_2 ton que são vendidos por R\$ 4,00 e R\$ 3,00 por tonelada, respectivamente. Na produção, são empregados três equipamentos diferentes. Para produzir 1 ton do bem 1 são necessárias 3 horas de uso do equipamento 1, 3 horas de uso do equipamento 2 e 4 horas de uso do equipamento 3. Para produzir 1 ton do bem 2 são necessárias 4 horas de uso do equipamento 1, 3 horas de uso do equipamento 2 e 2 horas de uso do equipamento 3. As disponibilidades diárias desses equipamentos são de 12 horas do equipamento 1, 10 horas do equipamento 2 e 8 horas do equipamento 3.

Formule o problema da maximização da receita da firma acima. Não é preciso resolver o problema.

Uma segunda firma quer alugar os equipamentos da primeira firma. Os aluguéis por hora dos respectivos equipamentos serão y_1 , y_2 e y_3 . Formule o problema de minimização de custos da segunda firma. Não é preciso resolver o problema.

6. Problema 3: o Problema de Transporte

Este problema foi, pela primeira vez, formulado pelo matemático Frank Lauren Hitchcock (1895-1957) no artigo “*The Distribution of a Product from Several Sources to Numerous Localities*”³², publicado em 1941.

O artigo de Hitchcock não utiliza nenhum resultado esotérico, ao contrário, a matemática empregada não ultrapassa a aritmética elementar. Sem dúvida, não foi o requinte que justificou a reputação do artigo. A fama justifica-se por ter sido o primeiro trabalho a formular matematicamente um problema na área de transportes, tema muito sensível durante períodos de guerra, e por propor um dispositivo simples, viável e criativo para achar uma solução que antecipa o chamado “Método Simplex”.

Exemplo 6: Digamos que uma firma tem dois depósitos a partir dos quais faz entregas para três lojas. Os gastos de transporte por unidade transportada de cada ponto de entrega para cada uma das lojas estão na tabela abaixo: c_{ij} é o custo de transportar uma unidade do depósito i para a loja j ; a última coluna indica quantas unidades estão disponíveis: no depósito um estão disponíveis 70 unidades e no depósito dois, 120 unidades; a última linha indica as solicitações: a loja 1 solicita 50 unidades, a loja dois, 60 unidades e a loja três, 80 unidades:

$$c_{11} = 2 \quad c_{12} = 5 \quad c_{13} = 1 \quad 70$$

$$c_{21} = 5 \quad c_{22} = 4 \quad c_{23} = 1 \quad 120$$

Se x_{ij} é o número de unidades transportadas do depósito i para a loja j , então o problema de minimização do custo de transporte é:

$$\begin{aligned} \min \quad & 2 \cdot x_{11} + 5 \cdot x_{12} + x_{13} + 5 \cdot x_{21} + 4 \cdot x_{22} + x_{23} \\ \text{sa} \quad & x_{11} + x_{21} = 50 \\ & x_{12} + x_{22} = 60 \\ & x_{13} + x_{23} = 80 \\ & x_{11} + x_{12} + x_{13} = 70 \\ & x_{21} + x_{22} + x_{23} = 120 \\ & x_{11} \geq 0, x_{12} \geq 0, x_{13} \geq 0, x_{21} \geq 0, x_{22} \geq 0, x_{23} \geq 0. \end{aligned}$$

Observe que, em relação às restrições, temos cinco equações. Uma das equações pode ser eliminada. Se somamos as três primeiras equações, temos o total de unidades demandadas. Se somamos as duas últimas equações, temos o total de unidades disponíveis, então

$$\text{Equação 1} + \text{Equação 2} + \text{Equação 3} - \text{Equação 4} - \text{Equação 5} = 0$$

Portanto, uma das linhas pode ser eliminada, pois é uma combinação linear das demais. Se escrevemos o sistema na forma matricial, temos:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 50 \\ 0 & 1 & 0 & 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 0 & 0 & 1 & 80 \\ 1 & 1 & 1 & 0 & 0 & 0 & 70 \\ 0 & 0 & 0 & 1 & 1 & 1 & 120 \end{pmatrix}$$

$x_{11} \quad x_{12} \quad x_{13} \quad x_{21} \quad x_{22} \quad x_{23} \mid b$

Por linha-redução obtemos:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 50 \\ 0 & 1 & 0 & 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 0 & 0 & 1 & 80 \\ 1 & 1 & 1 & 0 & 0 & 0 & 70 \\ 0 & 0 & 0 & 1 & 1 & 1 & 120 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 50 \\ 0 & 1 & 0 & 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 0 & 0 & 1 & 80 \\ 0 & 0 & 0 & -1 & -1 & -1 & -120 \\ 0 & 0 & 0 & 1 & 1 & 1 & 120 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 50 \\ 0 & 1 & 0 & 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 0 & 0 & 1 & 80 \\ 0 & 0 & 0 & -1 & -1 & -1 & -120 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & -1 & -70 \\ 0 & 1 & 0 & 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 0 & 0 & 1 & 80 \\ 0 & 0 & 0 & 1 & 1 & 1 & 120 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Facilmente verificamos que as linhas que restam são linearmente independentes. O posto da matriz é 4. Logo, temos quatro variáveis

dependentes, no caso $x_{11}, x_{12}, x_{13}, x_{21}$ e duas independentes x_{22} e x_{23} . O conjunto de soluções do sistema acima é um espaço afim de dimensão dois do $\mathbb{R}^6$:

$$\begin{aligned}x_{11} &= -70 + x_{22} + x_{23} \\x_{12} &= 60 - x_{22} \\x_{13} &= 80 - x_{23} \\x_{21} &= 120 - x_{22} - x_{23}.\end{aligned}$$

Obviamente,

$$\begin{aligned}x_{11} &= -70 + x_{22} + x_{23} \geq 0 \\x_{12} &= 60 - x_{22} \geq 0 \\x_{13} &= 80 - x_{23} \geq 0 \\x_{21} &= 120 - x_{22} - x_{23} \geq 0.\end{aligned}$$

O que implica que

$$\begin{aligned}x_{22} + x_{23} &\geq 70 \\x_{22} &\leq 60 \\x_{23} &\leq 80 \\x_{22} + x_{23} &\leq 120.\end{aligned}$$

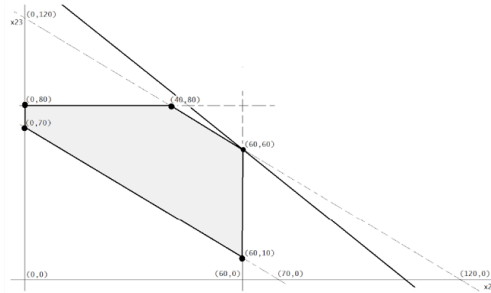
Se expressamos a função custo através das variáveis x_{22} e x_{23} , obtemos $C = 840 - 4 \cdot x_{22} - 3 \cdot x_{23}$ pois

$$C = 2 \cdot (-70 + x_{22} + x_{23}) + 5 \cdot (60 - x_{22}) + (80 - x_{23}) + 5 \cdot (120 - x_{22} - x_{23}) + 4 \cdot x_{22} + x_{33}.$$

Se aumentamos x_{22} e x_{23} diminuimos o custo, então, o problema se reduz a

$$\begin{aligned}\max \quad & 4 \cdot x_{22} + 3 \cdot x_{23} \\sa \quad & x_{22} + x_{23} \geq 70 \\& x_{22} \leq 60 \\& x_{23} \leq 80 \\& x_{22} + x_{23} \leq 120 \\& x_{22} \geq 0, x_{23} \geq 0.\end{aligned}$$

Veja o gráfico abaixo. Em cinza está a região determinada pelas restrições; a isoquanta mais alta de $4 \cdot x_{22} + 3 \cdot x_{23} = K$ é quando $K = 420$ e passa pelo ponto $(60, 60)$



Como $x_{22} = 60$ e $x_{23} = 60$, resulta em $x_{11} = 50$, $x_{12} = 0$, $x_{13} = 20$ $x_{21} = 0$ e custo mínimo de $C = 420$. Portanto, 50 unidades são levadas do depósito um para a loja um e a demanda da loja um está satisfeita; na loja dois são entregues 60 unidades a partir do depósito dois, as 20 unidades restantes no depósito um e as 60 unidades restantes no depósito dois são entregues na loja três.

Exercício 8

Suponhamos que temos dois depósitos e duas lojas. No primeiro depósito temos $x_{11} + x_{12} = 15 = o_1$; no segundo depósito temos $x_{21} + x_{22} = 30 = o_2$. Em relação às demandas: a primeira loja solicita $x_{11} + x_{21} = 20 = d_1$ e a segunda loja solicita $x_{12} + x_{22} = 25 = d_2$. Verifique a matriz desse problema e a correspondente linha-redução abaixo:

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 20 \\ 0 & 1 & 0 & 1 & 25 \\ 1 & 1 & 0 & 0 & 15 \\ 0 & 0 & 1 & 1 & 30 \end{array} \right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & -10 \\ 0 & 1 & 0 & 1 & 25 \\ 0 & 0 & 1 & 1 & 30 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right).$$

$x_{11} \quad x_{12} \quad x_{21} \quad x_{22} \quad b$

O posto da matriz é 3, que resultam em três variáveis dependentes e uma independente. A solução desse problema é uma reta no $\mathbb{R}^4$ que passa pelo ponto $(-10, 25, 30, 0)$ na direção do vetor $v' = (1, -1, -1, 1)$:

$$\begin{aligned} x_{11} &= -10 + x_{22} \\ x_{12} &= 25 - x_{22} \\ x_{21} &= 30 - x_{22}. \end{aligned}$$

Se $a_{11} = 5, a_{12} = 1, a_{21} = 3$ e $a_{22} = 5$ então $C = 65 + x_{22}$, qual é o menor valor possível para x_{22} ? Qual é o custo mínimo? Determine os valores x_{11} , x_{12} e x_{21} .

Exercício 9

Verifique por linha-redução

$$\left(\begin{array}{cccccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & d_3 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & o_1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & o_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & o_3 \end{array} \right) \sim \left(\begin{array}{cccccccc|c} 1 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & -o_1 - d_2 - d_3 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & o_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & (d_1 + d_2 + d_3) - (o_1 + o_2) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (d_1 + d_2 + d_3) - (o_1 + o_2 + o_3) \end{array} \right).$$

A estrutura acima é fixa para todos os problemas desse tipo. Digamos que temos m depósitos e n lojas, então as restrições são:

$$\begin{aligned} x_{11} + \dots + x_{1n} &= o_1 \\ &\vdots \\ x_{m1} + \dots + x_{mn} &= o_m \\ x_{11} + \dots + x_{m1} &= d_1 \\ &\vdots \\ x_{1n} + \dots + x_{mn} &= d_n. \end{aligned}$$

a matriz do problema é

$$\left(\begin{array}{cccc|cccc|c} 1 & \dots & 0 & 1 & \dots & 0 & \dots & 1 & \dots & 0 & d_1 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & \dots & 0 & \ddots & 0 & d_2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 1 & \dots & 0 & \dots & 1 & d_n \end{array} \right) \cdot \left(\begin{array}{ccc|c|c} I_{n \times n} & I_{n \times n} & \dots & I_{n \times n} & d \\ 1_{1 \times n} & 0 & \dots & 0 & o_1 \\ 0 & 1_{1 \times n} & \dots & 0 & o_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1_{1 \times n} & o_m \end{array} \right)$$

A condição de compatibilidade é $(o_1 + o_2 + \dots + o_m) = (d_1 + \dots + d_n)$, obviamente. Observe que são $n+m$ equações com $n \cdot m$ variáveis, que o posto da matriz é $n+m-1$, portanto, são $n+m-1$ variáveis dependentes e $n \cdot m - n - m + 1$ variáveis independentes.

6.1 A solução de Hitchcock

Vejamos como Hitchcock resolveu os problemas de transporte em seu artigo. Usando a notação do Hitchcock, retornamos ao primeiro

exemplo acima, dispondo as equações da maneira abaixo:

$$\begin{array}{ccc|c} x_{11} & x_{12} & x_{13} & 70 \\ x_{21} & x_{22} & x_{23} & 120 \\ \hline 50 & 60 & 80 & \end{array}$$

Se escolhermos $x_{11}=0$ e $x_{12}=0$, observando a tabela acima, resulta $x_{13}=70$, $x_{21}=50$, $x_{22}=60$ e $x_{23}=10$:

$$\begin{array}{ccc|c} 0 & 0 & x_{13} & 70 \\ x_{21} & x_{22} & x_{23} & 120 \\ \hline 50 & 60 & 10 & \end{array}$$

Encontramos a solução $x_{11}=0$, $x_{12}=0$, $x_{13}=70$, $x_{21}=50$, $x_{22}=60$ e $x_{23}=10$. São quatro equações independentes e seis variáveis, portanto, o sistema sempre resulta em duas variáveis independentes e quatro dependentes. Elegemos x_{11} e x_{12} como variáveis independentes. Se expressarmos as restrições através das variáveis independentes escolhidas resulta:

$$\begin{aligned} x_{13} &= 70 - x_{11} - x_{12} \\ x_{21} &= 50 - x_{11} \\ x_{22} &= 60 - x_{12} \\ x_{23} &= 80 - x_{13} = 80 - (70 - x_{11} - x_{12}) = 10 + x_{11} + x_{12}. \end{aligned}$$

A função de custo resulta em

$$\begin{aligned} C &= 2 \cdot x_{11} + 5 \cdot x_{12} + x_{13} + 5 \cdot x_{21} + 4 \cdot x_{22} + x_{23}, \\ C &= 2 \cdot x_{11} + 5 \cdot x_{12} + (70 - x_{11} - x_{12}) + 5 \cdot (50 - x_{11}) + 4 \cdot (60 - x_{12}) + (10 + x_{11} + x_{12}), \\ C &= 570 - 3 \cdot x_{11} + x_{12}. \end{aligned}$$

O custo está expresso em termos das variáveis independentes x_{11} e x_{12} . Observe que podemos diminuir o custo aumentando x_{11} . Então, usando a equação $x_{21}=50-x_{11}$, fazemos $x_{11}=50-x_{21}$, ou seja, x_{11} passa a ser variável dependente e x_{12} variável dependente. As equações acima se expressam como:

$$\begin{aligned} x_{13} &= 20 + x_{21} - x_{12} \\ x_{11} &= 50 - x_{21} \\ x_{22} &= 60 - x_{12} \\ x_{23} &= 60 - x_{21} + x_{12}. \end{aligned}$$

A função de custo torna-se $C = 420 + 3 \cdot x_{21} + x_{12}$. O valor mínimo para o custo é 420, quando $x_{21} = x_{12} = 0$. Os valores das outras variáveis resultam $x_{13} = 20$, $x_{11} = 50$, $x_{22} = 60$ e $x_{23} = 60$.

Exemplo 7: Estudemos o exemplo com o qual Hitchcock³³ ilustra o dispositivo que criou para a resolução dos problemas de transporte. Suponhamos que temos três pontos de distribuição e quatro pontos de recebimento, conforme a tabela abaixo:

x_{11}	x_{12}	x_{13}	x_{14}	25
x_{21}	x_{22}	x_{23}	x_{24}	25
x_{31}	x_{32}	x_{33}	x_{34}	50
15	20	30	35	

Digamos que os custos sejam

$$\begin{array}{llll} c_{11} = 10 & c_{12} = 5 & c_{13} = 6 & c_{14} = 7 \\ c_{21} = 8 & c_{22} = 2 & c_{23} = 7 & c_{24} = 6 \\ c_{31} = 9 & c_{32} = 3 & c_{33} = 4 & c_{34} = 8. \end{array}$$

As restrições são dadas por sete equações além das restrições $x_{ij} \geq 0$, já que todas as variáveis são não negativas.

Qualquer uma das equações é combinação linear das demais, pois a soma das equações $x_{11} + x_{12} + x_{13} + x_{14} = 25$, $x_{21} + x_{22} + x_{23} + x_{24} = 25$ e $x_{31} + x_{32} + x_{33} + x_{34} = 50$ que representa o total disponível é igual a soma das equações $x_{11} + x_{21} + x_{31} = 15$, $x_{12} + x_{22} + x_{32} = 20$, $x_{13} + x_{23} + x_{33} = 30$ e $x_{14} + x_{24} + x_{34} = 35$ que é o total demandado. Portanto, das sete equações, seis são linearmente independentes; como são 12 incógnitas, temos 6 variáveis dependentes e 6 variáveis independentes.

Se elegermos $x_{11} = 0$, $x_{12} = 0$ e $x_{13} = 0$, então $x_{14} = 25$. Se fizermos $x_{21} = 0$, $x_{22} = 0$ e $x_{24} = 0$, então $x_{23} = 25$, $x_{31} = 15$, $x_{32} = 20$, $x_{33} = 5$ e $x_{34} = 10$.

$$\begin{array}{cccc|cccc|cccc|c} 0 & 0 & 0 & 25 & 25 & 0 & 0 & 0 & 25 & 25 & 0 & 0 & 0 & 25 & 25 \\ x_{21} & x_{22} & x_{23} & x_{24} & 25 & 0 & 0 & 25 & 0 & 25 & 0 & 0 & 25 & 0 & 25 \\ x_{31} & x_{32} & x_{33} & x_{34} & 50 & x_{31} & x_{32} & x_{33} & x_{34} & 50 & 15 & 20 & 5 & 10 & 50 \\ \hline 15 & 20 & 30 & 10 & & 15 & 20 & 5 & 10 & & 15 & 20 & 30 & 35 & \end{array}$$

As variáveis independentes escolhidas são $x_{11} \geq 0$, $x_{12} \geq 0$, $x_{13} \geq 0$, $x_{21} \geq 0$, $x_{22} \geq 0$ e $x_{24} \geq 0$. Se fizermos as variáveis independentes iguais a zero, o último quadro acima determina o valor das variáveis dependentes. Em termos das variáveis independentes, as demais variáveis se expressam como:

$$\begin{aligned}x_{14} &= 25 - x_{11} - x_{12} - x_{13} \\x_{23} &= 25 - x_{21} - x_{22} - x_{24} \\x_{31} &= 15 - x_{11} - x_{21} \\x_{32} &= 20 - x_{12} - x_{22} \\x_{33} &= 5 - x_{13} + x_{21} + x_{22} + x_{24} \\x_{34} &= 10 + x_{11} + x_{12} + x_{13} - x_{24}.\end{aligned}$$

Para obter as equações acima utilizamos o primeiro quadro acima. O custo expresso nas variáveis independentes é:

$$\begin{aligned}C &= 10 \cdot x_{11} + 5 \cdot x_{12} + 6 \cdot x_{13} + 7 \cdot (25 - x_{11} - x_{12} - x_{13}) + 8 \cdot x_{21} + 2 \cdot x_{22} + 7 \cdot (25 - x_{21} - x_{22} - x_{24}) + \\&6 \cdot x_{24} + 9 \cdot (15 - x_{11} - x_{21}) + 3 \cdot (20 - x_{12} - x_{22}) + 4 \cdot (5 - x_{13} + x_{21} + x_{22} + x_{24}) + \\&8 \cdot (10 + x_{11} + x_{12} + x_{13} - x_{24})\end{aligned}$$

ou, ainda

$$C = 645 + 2 \cdot x_{11} + 3 \cdot x_{12} + 3 \cdot x_{13} - 4 \cdot x_{21} - 4 \cdot x_{22} - 5 \cdot x_{24}.$$

Se aumentarmos o valor de x_{21} , o valor de C diminui, então, na próxima etapa, vamos substituir x_{21} por outras variáveis, ou seja, passando de independente para dependente. No caso, vamos substituir a variável x_{21} por x_{31} usando a equação $x_{31} = 15 - x_{11} - x_{21}$, ou seja, x_{21} passa a ser uma variável dependente e x_{31} uma variável independente. O novo sistema é:

$$\begin{aligned}x_{14} &= 25 - x_{11} - x_{12} - x_{13} \\x_{23} &= 10 + x_{11} + x_{31} - x_{22} - x_{24} \\x_{21} &= 15 - x_{11} - x_{31} \\x_{32} &= 20 - x_{12} - x_{22} \\x_{33} &= 20 - x_{11} - x_{13} - x_{31} + x_{22} + x_{24} \\x_{34} &= 10 + x_{11} + x_{12} + x_{13} - x_{24}.\end{aligned}$$

As variáveis independentes são $x_{11} \geq 0$, $x_{12} \geq 0$, $x_{22} \geq 0$, $x_{13} \geq 0$, $x_{31} \geq 0$, e $x_{24} \geq 0$. A função de custo nestas variáveis resulta:

$$C = 585 + 6 \cdot x_{11} + 3 \cdot x_{12} + 3 \cdot x_{13} + 4 \cdot x_{31} - 4 \cdot x_{22} - 5 \cdot x_{24} .$$

Então, um resultado possível é $x_{11}=0$, $x_{12}=0$, $x_{13}=0$, $x_{22}=0$, $x_{31}=0$, e $x_{24}=0$. Nesse caso, o custo é $C=585$.

Podemos aumentar x_{22} . Troquemos a variável independente x_{22} pela variável dependente x_{23} usando a equação $x_{23}=10+x_{11}+x_{31}-x_{22}-x_{24}$. A nova variável independente é x_{23} , e x_{22} torna-se dependente. O sistema que resulta é:

$$\begin{aligned} x_{14} &= 25 - x_{11} - x_{12} - x_{13} \\ x_{22} &= 10 + x_{11} + x_{31} - x_{23} - x_{24} \\ x_{21} &= 15 - x_{11} - x_{31} \\ x_{32} &= 10 - x_{11} - x_{12} + x_{23} + x_{24} - x_{31} \\ x_{33} &= 20 - x_{13} - x_{23} \\ x_{34} &= 10 + x_{11} + x_{12} + x_{13} - x_{24} . \end{aligned}$$

As variáveis independentes são $x_{11} \geq 0$, $x_{12} \geq 0$, $x_{13} \geq 0$, $x_{23} \geq 0$, $x_{24} \geq 0$ e $x_{31} \geq 0$. A função de custo nessas variáveis resulta:

$$C = 545 + 2 \cdot x_{11} + 3 \cdot x_{12} + 3 \cdot x_{13} + 4 \cdot x_{23} - x_{24} .$$

Então, um resultado possível é $x_{11}=0$, $x_{12}=0$, $x_{13}=0$, $x_{23}=0$, $x_{24}=0$ e $x_{31}=0$. Nesse caso, o custo é $C=545$.

Finalmente, vamos trocar a variável x_{24} pela variável x_{34} :

$$\begin{aligned} x_{14} &= 25 - x_{11} - x_{12} - x_{13} \\ x_{22} &= -x_{12} - x_{13} - x_{23} + x_{31} + x_2 \\ x_{21} &= 15 - x_{11} - x_{31} \\ x_{32} &= 20 + x_{13} + x_{23} - x_{31} - x_{34} \\ x_{33} &= 30 - x_{13} - x_{23} \\ x_{24} &= 10 + x_{11} + x_{12} + x_{13} - x_{34} . \end{aligned}$$

As variáveis independentes são $x_{11} \geq 0$, $x_{12} \geq 0$, $x_{13} \geq 0$, $x_{23} \geq 0$, $x_{31} \geq 0$ e $x_{34} \geq 0$. A função de custo nessas variáveis se expressa como:

$$C = 535 + x_{11} + 2 \cdot x_{12} + 2 \cdot x_{13} + 4 \cdot x_{23} + x_{34} .$$

Claramente, resulta que o menor valor para o custo é $C=535$, obtido quando $x_{11}=0$, $x_{12}=0$, $x_{13}=0$, $x_{23}=0$, $x_{31}=0$ e $x_{34}=0$. Portanto, obtivemos uma solução:

os valores das variáveis dependentes são $x_{14} = 25$, $x_{22} = 0$, $x_{21} = 15$, $x_{32} = 20$, $x_{33} = 30$ e $x_{24} = 10$

.

Duas ideias fundamentam o método de Hitchcock. A primeira é que em cada etapa, entre as 12 variáveis, são escolhidas 6 variáveis para serem variáveis independentes e as restantes são dependentes das 6 escolhidas. A seguir, as variáveis dependentes e a função objetivo são expressas através das variáveis independentes. No caso, a função objetivo é a função custo.

A segunda ideia é que as alterações dos conjuntos das variáveis independentes em cada etapa têm por objetivo reduzir o número de variáveis com coeficientes negativos na expressão da função objetivo.

Se através dos dois procedimentos acima obtemos uma expressão da função de custo através de variáveis independentes apenas com coeficientes não negativos, então estamos num ponto de mínimo da função de custo, pois se aumentarmos o valor de alguma das variáveis, obviamente o custo aumenta.

Os três parágrafos anteriores resumem o que é o método simplex e, por essa razão, muito justamente, George Dantzig³⁴, considerado o pai do método simplex, e Tjalling Koopmans³⁵ fazem referências ao artigo de Hitchcock.

A ausência da variável independente x_{31} na expressão final da função de custo sugere que a solução obtida acima não é única. Se examinamos as soluções:

$$\begin{aligned}x_{14} &= 25 - x_{11} - x_{12} - x_{13} \\x_{22} &= -x_{12} - x_{13} - x_{23} + x_{31} + x_3 \\x_{21} &= 15 - x_{11} - x_{31} \\x_{32} &= 20 + x_{13} + x_{23} - x_{31} - x_{34} \\x_{33} &= 30 - x_{13} - x_{23} \\x_{24} &= 10 + x_{11} + x_{12} + x_{13} - x_{34},\end{aligned}$$

Observamos que x_{31} pode variar de maneira que $0 \leq x_{31} \leq 15$. Se x_{31} varia, as variáveis x_{21} , x_{22} e x_{32} também variam:

$$\begin{pmatrix} x_{21} \\ x_{22} \\ x_{31} \\ x_{32} \end{pmatrix} = \begin{pmatrix} 15 \\ 0 \\ 0 \\ 20 \end{pmatrix} + \begin{pmatrix} -15 \\ 15 \\ 15 \\ -15 \end{pmatrix} \cdot t, 0 \leq t \leq 1$$

Entretanto, o custo não se altera e permanece $C = 535$, pois as variáveis x_{21} , x_{22} , x_{32} e x_{31} não estão na expressão $C = 535 + x_{11} + 2 \cdot x_{12} + 2 \cdot x_{13} + 4 \cdot x_{23} + x_{34}$ do custo! Portanto, acabamos de descobrir uma aresta do poliedro determinado pelas restrições do problema que satisfaz o problema da minimização do custo! Observe que esse poliedro está no $\mathbb{R}^{12}$! Pode parecer um pouco exagerado, mas por uma única vez, vale a pena expressar essa aresta:

$$\begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{24} \\ x_{31} \\ x_{32} \\ x_{33} \\ x_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 25 \\ 15 \\ 0 \\ 0 \\ 10 \\ 0 \\ 20 \\ 30 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -15 \\ 15 \\ 0 \\ 0 \\ 15 \\ -15 \\ 0 \\ 0 \end{pmatrix} \cdot t, 0 \leq t \leq 1$$

Exercício 10

A aresta acima está sobre uma reta do $\mathbb{R}^{12}$. Indique um vetor que esteja na direção dessa reta. Determine um hiperplano paralelo ao hiperplano $C = 10 \cdot x_{11} + 5 \cdot x_{12} + 6 \cdot x_{13} + 7 \cdot x_{14} + 8 \cdot x_{21} + 2 \cdot x_{22} + 7 \cdot x_{23} + 6 \cdot x_{24} + 9 \cdot x_{31} + 3 \cdot x_{32} + 4 \cdot x_{34} + 8 \cdot x_{41}$

tal que o poliedro determinado pelas restrições esteja inteiramente num dos lados do plano.

O processo elaborado por Hitchcock implica que, no sistema de $n+m-1$ equações linearmente independentes, que expressa a restrição (existem ainda as inequações $x_v \geq 0$), realizemos certo número de

linhas-reduções nas quais expressamos $n+m-1$ variáveis dependentes através de $n+m-(n+m-1)$ variáveis independentes. No método simplex também temos que empregar esse procedimento. Então, vejamos primeiramente como podemos alterar o conjunto de variáveis dependentes e independentes.

Observe que o custo está expresso em termos das variáveis independentes $25 - x_{11} - x_{12} - x_{13}$, $25 - x_{21} - x_{22} - x_{24}$.

Exercício 11

Suponhamos que temos três pontos de distribuição e três pontos de recebimento conforme a tabela abaixo:

x_{11}	x_{12}	x_{13}	20
x_{21}	x_{22}	x_{23}	15
x_{31}	x_{32}	x_{33}	10
5	20	20	

Vamos determinar duas variáveis independentes: na terceira linha, escolhamos as variáveis x_{31} e x_{32} como independentes e x_{33} como dependente:

x_{11}	x_{12}	x_{13}	20
x_{21}	x_{22}	x_{23}	15
0	0	10	10
5	20	10	

Na segunda linha, escolhamos x_{21} e x_{23} como independentes e x_{22} como dependente:

x_{11}	x_{12}	x_{13}	20	5	5	10	20	
0	15	0	15	→	0	15	0	15
0	0	10	10		0	0	10	10
5	5	10			5	20	20	

As variáveis independentes são x_{21}, x_{23}, x_{31} e x_{32} , e as demais $x_{11}, x_{12}, x_{13}, x_{22}$ e x_{33} . São seis equações e nove variáveis, qualquer uma das equações é soma das demais e quaisquer cinco equações são linearmente

independentes. O sistema que expressa as restrições de igualdades do problema:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{31} \\ x_{32} \\ x_{33} \end{pmatrix} = \begin{pmatrix} 20 \\ 15 \\ 10 \\ 5 \\ 20 \\ 20 \end{pmatrix}$$

É fácil perceber que

$$\text{Equação 1} + \text{Equação 2} + \text{Equação 3} = \text{Equação 4} + \text{Equação 5} + \text{Equação 6},$$

o total distribuído é igual ao total recebido. Se retirarmos uma das equações, as demais são linearmente independentes. Logo, o posto da matriz é 5, então podemos expressar cinco das variáveis em função das quatro restantes. Por exemplo, acima escolhemos expressar x_{11} , x_{12} , x_{13} , x_{22} e x_{33} através de x_{21} , x_{23} , x_{31} e x_{32} o que resulta em:

$$\begin{aligned} x_{11} &= 5 - x_{21} - x_{31} \\ x_{12} &= 20 - x_{22} - x_{32} = 20 - (15 - x_{21} - x_{23}) - x_{32} = 5 + x_{21} + x_{23} - x_{32} \\ x_{13} &= 20 - x_{23} - (10 - x_{31} - x_{32}) = 10 - x_{23} + x_{31} + x_{32} \\ x_{22} &= 15 - x_{21} - x_{23} \\ x_{33} &= 10 - x_{31} - x_{32}. \end{aligned}$$

Se os custos de transporte são

$C_{11} = 3$	$C_{12} = 4$	$C_{13} = 5$
$\times$	$C_{22} = 3$	$\times$
$\times$	$C_{32} = 3$	$C_{33} = 6$

determine a distribuição que implica o menor custo. Observe que o depósito dois só pode fazer entregas para loja dois, e que para o depósito três é inviável fazer entregas na loja um.

Exercício 12

A planta 1 produz 15 toneladas e a planta 2, 8 toneladas por semana. O ponto 1 recebe 5 toneladas e o ponto 2 recebe 18 toneladas. Os custos de transporte em reais por toneladas estão na tabela abaixo:

	Ponto 1	Ponto 2
Firma 1	C_{11}	C_{12}
Firma 2	C_{21}	C_{22}

Se supomos que $C_{12} + C_{21} > C_{11} + C_{22}$, determine como deve ser a distribuição de maneira a minimizar o custo de transporte. Vamos notar o produzido na firma i e transportado para o ponto j por x_{ij} .

7. O problema C1 de Leonid Kantorovich

Exemplo 8: Digamos que temos três maneiras de programar uma máquina. Em cada uma das maneiras podemos fabricar certo número de peças do tipo A e do tipo B. No programa 1, podemos fabricar 10 peças do tipo A e 4 peças do tipo B; no programa 2, fabricamos 4 peças do tipo A e 6 peças do tipo B; no programa 3, fabricamos 5 peças do tipo A e 3 peças do tipo B. Isso durante o período unitário de trabalho $H=1$ (hora, dias, mês, etc.). Veja a matriz abaixo:

	Peças do tipo A		Peças do tipo B
Programa 1	10/H	e	4/H
ou			
Programa 2	4/ H	e	6/H
ou			
Programa 3	5/ H	e	3/H
ou			

Seja $h' = (h_1 \ h_2 \ h_3)$, $H = h_1 + h_2 + h_3 = 1$, $h_i, i = 1, 2, 3$ a representação da fração do tempo reservada a cada uma das técnicas, respectivamente. Consideremos a matriz:

$$C = \begin{pmatrix} 10 & 4 \\ 4 & 6 \\ 5 & 3 \end{pmatrix},$$

Então

$$h' \cdot C = (h_1 \ h_2 \ h_3) \cdot \begin{pmatrix} 10 & 4 \\ 4 & 6 \\ 5 & 3 \end{pmatrix} = (10 \cdot h_1 + 4 \cdot h_2 + 5 \cdot h_3 \quad 4 \cdot h_1 + 6 \cdot h_2 + 3 \cdot h_3) = (z_1 \ z_2)$$

em que $10 \cdot h_1 + 4 \cdot h_2 + 5 \cdot h_3 = z_1$ é o número de peças fabricadas do tipo A, e $4 \cdot h_1 + 6 \cdot h_2 + 3 \cdot h_3 = z_2$ de peças fabricadas do tipo B, respectivamente.

Essas peças são articuladas num conjunto maior, por exemplo, numa máquina que necessita de uma peça de cada tipo. Portanto, o problema é encerrar o período de trabalho com o número máximo de conjuntos $\{A, B\}$ e com um mínimo de peças do tipo A ou B em excesso. Os objetivos são dois:

- i. igualar $z = z_1 = z_2 = 10 \cdot h_1 + 4 \cdot h_2 + 5 \cdot h_3 = 4 \cdot h_1 + 6 \cdot h_2 + 3 \cdot h_3$ e
- ii. maximizar z

Para satisfazer o primeiro objetivo, observemos que:

$$\begin{aligned} z_1 - z_2 &= 6 \cdot h_1 - 2 \cdot h_2 + 2 \cdot h_3 = 0 \\ h_1 + h_2 + h_3 &= 1, \end{aligned}$$

Portanto, estamos buscando a reta que está na intersecção destes planos, resultando:

$$\begin{aligned} h_1 &= \frac{1}{4} - \frac{1}{2} \cdot t \\ h_2 &= \frac{3}{4} - t \\ h_3 &= t. \end{aligned}$$

Como $h_1, h_2, h_3 \geq 0$, então $0 \leq t \leq \frac{1}{2}$. Variando o parâmetro t , obtemos as distribuições de tempo entre os processos que geram o mesmo

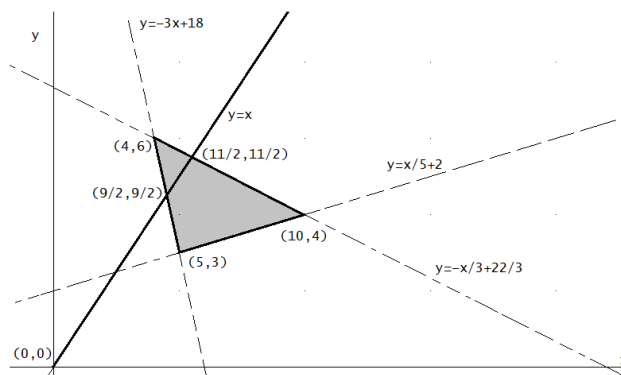
número z de peças do tipo A e do tipo B.

Para satisfazer o segundo objetivo, observamos que

$$z = 4 \cdot h_1 + 6 \cdot h_2 + 3 \cdot h_3 = 4 \cdot \left(\frac{1}{4} - \frac{1}{2} \cdot t \right) + 6 \cdot \left(\frac{3}{4} - t \right) + 3 \cdot t = \frac{11}{2} - 5 \cdot t$$

Como queremos maximizar z , então $t=0$, resultando em $z=\frac{11}{2}$ conjuntos. Fabricaremos ao longo do período $\frac{11}{2}=5,5$ tanto de peças do tipo A quanto do tipo B. Obviamente, essa solução não sendo um valor inteiro é aceitável em algumas situações e em outras não. Para corrigir esse defeito, temos que melhorar nosso método, mas, por enquanto, deixemos assim. Em relação às frações de tempo dedicadas a cada um dos processos, temos $h_1=\frac{1}{4}$, $h_2=\frac{3}{4}$ e $h_3=0$.

Vamos ver outra solução para o mesmo problema. Se empregarmos apenas o processo 1, ao longo da jornada de trabalho fabricamos 10 peças do tipo A e 4 peças do tipo B, da mesma maneira, em relação aos outros processos conforme a tabela acima. O gráfico abaixo representa o triângulo que contém todas as possibilidades de fabricação de peças do tipo A e do tipo B através das combinações dos processos:



Mais à frente, caracterizaremos esse triângulo como o conjunto das combinações convexas dos vértices $(5,3)$, $(4,6)$ e $(10,4)$. Observando no gráfico, facilmente notamos que os conjuntos possíveis com igual número de peças do tipo A e do tipo B se situam no segmento de

reta entre os pontos $\left(\frac{9}{2}, \frac{9}{2}\right)$ e $\left(\frac{11}{2}, \frac{11}{2}\right)$, ou seja, na intersecção da reta $y=x$ com o triângulo. Claramente, a solução está em $\left(\frac{11}{2}, \frac{11}{2}\right)$ na intersecção das retas $y=x$ e $y=-\frac{x}{3}+\frac{22}{3}$ que unem os pontos $(4,6)$ e $(10,4)$. Portanto, estamos utilizando os processos 1 e 2, cujos extremos são $(4,6)$ e $(10,4)$; o processo 3 não é utilizado, então $h_3=0$. A solução $\left(\frac{11}{2}, \frac{11}{2}\right)$ e $(10,4)$ está entre $(4,6)$ e $(10,4)$, ou seja, $\left(\frac{11}{2}, \frac{11}{2}\right)=h_1 \cdot (10,4)+h_2 \cdot (4,6)$, que resulta $h_1=\frac{1}{4}$ e $h_2=\frac{3}{4}$.

Mais uma observação importante: a solução está na intersecção das retas $y=x$ e $y=-\frac{x}{3}+\frac{22}{3}$. Na reta $y=-\frac{x}{3}+\frac{22}{3}$ se deixamos de fabricar três peças do tipo A, em compensação, fabricamos apenas uma peça do tipo B num mesmo período de tempo, digamos τ . Para igualar os números de peças dos dois tipos, devemos dedicar $h_1=\frac{1}{4} \cdot T$ e $h_2=\frac{3}{4} \cdot T$.

A declividade da reta $y=-\frac{x}{3}+\frac{22}{3}$ é $-\frac{1}{3}$, nós tomamos a fração $\frac{1}{3}$ como a razão entre as frações de tempo $\frac{h_1}{h_2}=\frac{1}{3}$. Kantorovich chamou estes números h_1 e h_2 de **multiplicadores** e sua versão da Programação Linear de “**Método dos Multiplicadores**”.

Exercício 13

Digamos que os conjuntos devam ser formados por duas peças do tipo A e uma peça do tipo B: $\{A, A, B\}$. Descubra o número máximo de conjuntos produzidos no período de trabalho e como o tempo é distribuído entre os processos. Resposta: são produzidos $\frac{22}{5}$ conjuntos, $h_1=\frac{4}{5}$ e $h_2=\frac{1}{5}$.

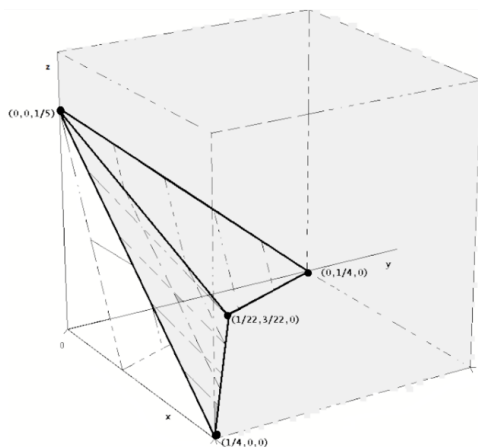
Vamos apresentar ainda outra solução para o mesmo problema. Já que z_1 é o número de peças do tipo A fabricadas e, da mesma forma, z_2 é o número de peças do tipo B, observe que $z=\min(z_1, z_2)$ é o número de conjuntos $\{A, B\}$ fabricados. O nosso problema é escolher (h_1, h_2, h_3) de maneira a maximizar o número de conjuntos $z=\min(z_1, z_2)$.

Observe que $10 \cdot h_1 + 4 \cdot h_2 + 5 \cdot h_3 \geq z$ e $4 \cdot h_1 + 6 \cdot h_2 + 3 \cdot h_3 \geq z$, portanto, $10 \cdot \left(\frac{h_1}{z}\right) + 4 \cdot \left(\frac{h_2}{z}\right) + 5 \cdot \left(\frac{h_3}{z}\right) \geq 1$ e $4 \cdot \left(\frac{h_1}{z}\right) + 6 \cdot \left(\frac{h_2}{z}\right) + 3 \cdot \left(\frac{h_3}{z}\right) \geq 1$. Substituindo $y_i = \frac{h_i}{z}, i=1, 2, 3$ obtemos $10 \cdot y_1 + 4 \cdot y_2 + 5 \cdot y_3 \geq 1$ e $4 \cdot y_1 + 6 \cdot y_2 + 3 \cdot y_3 \geq 1$.

Por outro lado, $y_1 + y_2 + y_3 = \frac{h_1 + h_2 + h_3}{z} = \frac{1}{z}$. Queremos maximizar z , que é o mesmo que minimizar $y_1 + y_2 + y_3 = \frac{1}{z}$. Portanto, nosso problema é:

$$\begin{aligned} \min \quad & y_1 + y_2 + y_3 \\ \text{sa} \quad & 10 \cdot y_1 + 4 \cdot y_2 + 5 \cdot y_3 \geq 1 \\ & 4 \cdot y_1 + 6 \cdot y_2 + 3 \cdot y_3 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{aligned}$$

Abaixo, a região viável está destacada junto com os correspondentes vértices: $(0, 0, \frac{1}{5})$, $(\frac{1}{4}, 0, 0)$, $(\frac{1}{22}, \frac{3}{22}, 0)$ e $(0, \frac{1}{4}, 0)$.



O mínimo de $y_1 + y_2 + y_3 = \frac{1}{z}$ está no vértice $(\frac{1}{22}, \frac{3}{22}, 0)$, em que $y_1 = \frac{1}{22}$, $y_2 = \frac{3}{22}$ e $y_3 = 0$ e, portanto, $\frac{1}{z} = \frac{2}{11}$, ou seja, o número máximo possível de conjuntos produzidos num período de trabalho é $\frac{11}{2} = 5,5$. Para isso, devemos empregar $h_1 = \frac{1}{22} \cdot \frac{11}{2} = \frac{1}{4}$, $h_2 = \frac{3}{22} \cdot \frac{11}{2} = \frac{3}{4}$ e $h_3 = 0 \cdot \frac{11}{2} = 0$, ou seja, 25% do período de tempo a máquina utiliza o programa 1 e 75% do período, o programa 2. Ao longo do período de trabalho, fabricaremos $\frac{11}{2} = 5,5$ tanto de peças do tipo A quanto do tipo B.

Essa última solução do problema C1 é importante para as soluções dos problemas que veremos a seguir.

Vejamos como o Kantorovich se refere aos problemas dos quais C1 é um exemplo:

quando a produção visa certo número de conjuntos que não estejam necessariamente completos, os componentes que faltam talvez tenham que serem comprados a um preço

mais alto ou, se alternativamente, certos componentes sejam produzidos em excesso talvez tenham que ser vendidos a um preço mais baixo. Em tais casos a quantidade real de conjuntos completos é muito importante para a determinação dos custos de produção. (“**Mathematical Methods of Production, Planning and Organization**”, 1939, p. 230)³⁶

No ensaio, Kantorovich diz que o problema que o motivou foi apresentado ao Instituto de Matemática e Mecânica da Universidade de Leningrado (S. Petersburgo) pelo Laboratório de Madeira Compensada. Segundo Roy Gardner, no artigo “*L. V. Kantorovich: The Price Implications of Optimal Planning*”³⁷, o problema se relacionava ao uso de tornos para retirar as cascas de troncos de diferentes tipos de árvores. Pois bem, a análise da questão revelou um conjunto de problemas referentes à organização produtiva que conduzem ao mesmo problema matemático:

estas questões se relacionam à distribuição ótima do emprego de máquinas-ferramentas e dispositivos, redução máxima do desperdício, melhor utilização do material bruto e dos materiais disponíveis no local, combustível, transporte, etc. Estes problemas não são adequados de maneira imediata para a análise matemática; ou melhor, parecem superficialmente apropriados e mesmo revelam-se simples, mas o método de solução é absolutamente impraticável dado que envolvem a solução de dezenas de milhares de equações ou mesmo milhões de sistemas de equações. Nós tivemos sucesso em encontrar um método relativamente simples para resolver este grupo de problemas que é aplicável a todos os problemas mencionados acima e que é suficientemente direto e efetivo para a sua aplicação às condições práticas.” (“**Mathematical Methods of Production, Planning and Organization**”, 1939, p. 226)³⁸

O método ao qual Kantorovich se refere é chamado “método dos multiplicadores” e é essencialmente a descoberta da “Programação Linear”.

O chamado **problema A** de Kantorovich é o seguinte: suponhamos que temos $i = 1, 2, \dots, n$ máquinas que podem produzir $k = 1, 2, \dots, m$ componentes diferentes. Uma máquina do tipo i pode produzir α_{ik} componentes do tipo k por período de funcionamento $H = 1$. O problema é determinar o número de peças de cada tipo a ser produzida por máquina, de forma a maximizar o número de produtos acabados. Digamos que h_k

é fração do tempo de funcionamento $H=1$ reservado pela máquina i para produzir componentes do tipo k , então $\sum_{k=1}^m k_{ik} = H=1$, portanto, o número de componentes do tipo k é $z_k = \sum_{i=1}^n \alpha_{ik} \cdot h_{ik}$ no período H . Assim sendo, se necessitamos em cada conjunto de componentes um componente de cada tipo, então o problema é:

$$\begin{array}{ll} \max & z_k \\ \text{sa} & z_1 = z_2 = \dots = z_m. \end{array}$$

O **problema C** que Kantorovich apresentou é o seguinte: suponhamos que temos $i=1,2,\dots,n$ máquinas que produzem $k=1,2,\dots,m$ componentes diferentes. Digamos que temos $l=1,2,\dots,p$ programações diferentes para cada máquina. Uma máquina do tipo i pode produzir γ_{ikl} componentes do tipo k se utilizar a programação, ou técnica, l por período de funcionamento $H=1$. Se h_{il} é fração do tempo de funcionamento $H=1$ reservado pela máquina i através da técnica l , então $\sum_{l=1}^p h_{il} = H=1$. Desse modo, o número de componentes do tipo k é $z_k = \sum_{i=1}^n \sum_{l=1}^p \gamma_{ikl} \cdot h_{il}$ no período H . Portanto, se necessitamos em cada conjunto de componentes de um componente de cada tipo, então o problema é:

$$\begin{array}{ll} \max & z_k \\ \text{sa} & z_1 = z_2 = \dots = z_m. \end{array}$$

O problema que tratamos acima é o chamado **problema C1**, quando temos uma máquina e podemos fracionar o período de trabalho $H=1$ entre as diferentes técnicas $\sum_{l=1}^p h_{il} = H=1$. Nesse caso, $z_k = \sum_{l=1}^p \gamma_{ikl} \cdot h_{il}$ o número de componentes do tipo k , em que γ_{ikl} é o número de componentes k que a máquina pode produzir usando a técnica l no período H . Novamente, o problema é:

$$\begin{array}{ll} \max & z_k \\ \text{sa} & z_1 = z_2 = \dots = z_m. \end{array}$$

Exercício 14

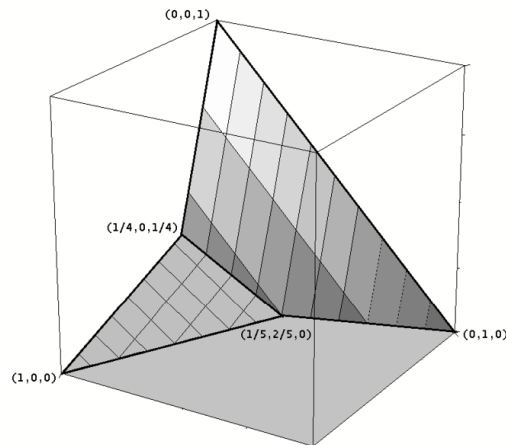
Digamos que temos três maneiras de programar uma máquina. Em cada uma das maneiras, podemos fabricar certo número de peças do tipo A e do tipo B. No programa 1, podemos fabricar 3 peças do tipo A e 1 peça do tipo B; no programa 2, fabricamos 1 peça do tipo A e 2 peças do tipo B; no programa 3, fabricamos 1 peça do tipo A e 3 peças do tipo B. Isso durante o período unitário de trabalho $H=1$:

$$C = \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

Formule o problema para a maximização dos conjuntos $\{A,B\}$ e determine o número de conjuntos produzidos:

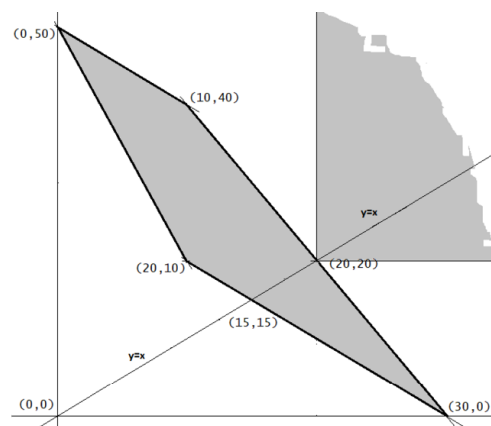
$$\begin{array}{ll} \min & y_1 + y_2 + y_3 \\ \text{sa} & 3 \cdot y_1 + y_2 + y_3 \geq 1 \\ & y_1 + 2 \cdot y_2 + 3 \cdot y_3 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

Use o gráfico abaixo:



Exemplo 9: No artigo “L. V. Kantorovich: *The Price Implications of Optimal Planning*” (1990), Roy Gardner fornece um exemplo muito esclarecedor sobre o método empregado por Kantorovich. Na verdade, já usamos este procedimento para resolver nosso primeiro exemplo do problema C1. Suponhamos que temos dois tornos para

descascar troncos de madeira e dois tipos de madeiras: o primeiro torno no período de tempo $H=1$ pode descascar 10 troncos do tipo 1 de madeira ou 10 troncos do tipo 2; o segundo torno pode descascar 20 troncos do primeiro tipo de madeira ou 40 troncos do tipo 2.



	Madeira Tipo 1		Madeira Tipo 2
Máquina 1	10	ou	10
Máquina 2	20	ou	40

No desenho acima, o quadrilátero com os vértices $(30,0)$, $(10,20)$, $(0,50)$ e $(0,0)$ delimita a região com as combinações convexas desses quatro vértices que definem as possibilidades de produção, por exemplo, se um período de tempo H for utilizado inteiramente para descascar madeiras do tipo 1, no fim do período teremos 30 troncos descascados do tipo 1 e nenhum do tipo 2; se empregarmos a máquina 1 durante todo o período para descascar madeira do tipo 2 e a máquina 2 para descascar madeiras do tipo 1, obtemos o vértice $(10,20)$.

Se o objetivo é obter o mesmo número de troncos descascados dos dois tipos de madeira, facilmente vemos que a resposta está no ponto $(10,20)$ na interseção da reta $y = -2 \cdot x + 60$ e $y = x$, como fizemos anteriormente. Digamos que as frações de tempo que cada uma das

máquinas precisa dedicar para cada tipo de madeira são fornecidas pelas equações abaixo:

$$\begin{aligned} 10 \cdot h_{11} + 20 \cdot h_{21} &= 20 \\ 10 \cdot h_{12} + 40 \cdot h_{22} &= 20 \\ h_{11} + h_{12} &= 1 \\ h_{21} + h_{22} &= 1, \end{aligned}$$

em que h_{ij} é a fração de tempo empregada pela máquina i na madeira do tipo j . A solução é $h_{11} = 1$, $h_{12} = 0$ e $h_{21} = h_{22} = \frac{1}{2}$.

Podemos buscar a solução como fizemos anteriormente. Considere:

$$(h_{11} \quad h_{12} \quad h_{21} \quad h_{22}) \cdot \begin{pmatrix} 10 & 0 \\ 0 & 10 \\ 20 & 0 \\ 0 & 40 \end{pmatrix} \begin{array}{l} \leftarrow \text{programa 1 na máquina 1} \\ \leftarrow \text{programa 2 na máquina 1} \\ \leftarrow \text{programa 1 na máquina 2} \\ \leftarrow \text{programa 2 na máquina 2.} \end{array}$$

Como

$$(h_{11} \quad h_{12} \quad h_{21} \quad h_{22}) \cdot \begin{pmatrix} 10 & 0 \\ 0 & 10 \\ 20 & 0 \\ 0 & 40 \end{pmatrix} = (10 \cdot h_{11} + 20 \cdot h_{21} \quad 10 \cdot h_{12} + 40 \cdot h_{22}) = (z_1 \quad z_2)$$

então $10 \cdot h_{11} + 20 \cdot h_{21} = z_1 \geq z = \min(z_1, z_2)$ e $10 \cdot h_{12} + 40 \cdot h_{22} = z_2 \geq z$. Logo, se $x_{ij} = \frac{h_{ij}}{z}$, então $10 \cdot x_{11} + 20 \cdot x_{21} \geq 1$ e $10 \cdot x_{12} + 40 \cdot x_{22} \geq 1$; como $h_{11} + h_{12} = 1$ e $h_{21} + h_{22} = 1$, então $x_{11} + x_{12} = \frac{1}{z} = x_{21} + x_{22}$ e, portanto, $x_{11} + x_{12} - x_{21} - x_{22} = 0$. Por outro lado, $x_{11} + x_{12} + x_{21} + x_{22} = \frac{2}{z}$. Como o objetivo é maximizar z , então o objetivo é minimizar $\frac{2}{z}$, portanto, podemos formular o seguinte problema para determinar os h_{ij} :

$$\begin{array}{ll} \min & x_{11} + x_{12} + x_{21} + x_{22} \\ \text{sa} & 10 \cdot x_{11} + 20 \cdot x_{21} \geq 1 \\ & 10 \cdot x_{12} + 40 \cdot x_{22} \geq 1 \\ & x_{11} + x_{12} - x_{21} - x_{22} \geq 0 \\ & x_{11} \geq 0, x_{12} \geq 0, x_{21} \geq 0, x_{22} \geq 0, \end{array}$$

que é semelhante aos problemas que já resolvemos e cuja solução obtemos pelo método dos multiplicadores do Kantorovich.

7.1 Leonid Vitalevich Kantorovich

Leonid Vitalevich Kantorovich (1912-1986) estudou na Universidade Estatal de Leningrado (S. Petersburgo), doutorando-se, em 1930, em Matemática aos 18 anos. Assim como Émile Borel e John von Neumann – a quem estamos constantemente fazendo referências –, desde criança foi visto como um prodígio.

Como Roy Gardner assinalou no artigo “*L. V. Kantorovich: The Price Implications of Optimal Planning*” (1990), as trajetórias de vida de Kantorovich e von Neumann têm muitos pontos em comum: ambos são intelectualmente excepcionais; muito jovens publicaram trabalhos fundamentais para a Matemática do século 20; em algum momento de suas vidas, dedicam-se à Matemática Aplicada, que acaba por dominar quase inteiramente seus interesses; ambos se sensibilizaram em relação à ameaça nazista e disponibilizaram suas habilidades e aptidões para os esforços da guerra. Na verdade, Kantorovich, Borel e von Neumann tiveram vidas semelhantes, como veremos mais a frente.

Após o doutorado, Kantorovich lecionou na Universidade Estatal de Leningrado até 1960; de 1961 até 1976, lecionou Economia Matemática na sucursal da Academia de Ciências da União Soviética na Sibéria e foi responsável pelo surgimento de uma escola de planejamento econômico voltado para a Economia soviética; de 1971 até 1986, Kantorovich foi diretor de pesquisa do Instituto de Planejamento Econômico, em Moscou. Nesse último período, segundo Gardner, Kantorovich desfrutou de maior influência na política econômica.

A versão de Kantorovich para a Programação Linear decorreu de uma consulta feita pelos engenheiros do Laboratório de Madeira Compensada ao Instituto de Matemática e Mecânica da

Universidade de Leningrado em 1937. Os engenheiros queriam cortar chapas produzindo um conjunto determinado de componentes diferentes com um mínimo de desperdício, segundo Gardner (p. 639) ou ainda, conforme Gardner (p. 642), determinar a maneira ótima de usar um conjunto de tornos que retiram as cascas de troncos de diferentes tipos de árvores de forma a produzir conjuntos determinados de diferentes troncos. A resposta de Kantorovich foi o ensaio *“Mathematical Methods of Production, Planning and Organization”* (em russo), publicado em 1939.

Kantorovich, para resolver esses problemas, utilizou pela primeira vez – antes de von Neumann e de Jean de Ville (1910-1989), um matemático ligado a Borel – um conjunto de resultados da Análise Convexa chamados *“Teoremas de Separação”*, que se tornaram fundamentais para a Teoria da Otimização. No começo da década de 1940, apenas dois matemáticos detinham o uso desses dispositivos para uso na Teoria da Otimização: von Neumann e Kantorovich. Gardner³⁹ faz uma comparação muito ilustrativa: Samuelson afirma que a primeira versão do seu livro *“Foundations of Economic Analysis”* foi escrita em 1937 e que a primeira versão conhecida é de 1941; nesse livro, o instrumento teórico empregado para otimização é o velho método de Lagrange, que vem do século 18.

Os trabalhos de Kantorovich foram conhecidos no Ocidente a partir de 1959, graças aos esforços de Tjalling Koopmans. O ensaio de Kantorovich foi traduzido e publicado no periódico *Management Science* em 1960. Anteriormente, em 1958, no mesmo periódico, havia sido publicado um outro artigo de Kantorovich, *“On the Translocation of Masses”*, escrito em 1942, sobre o problema de transporte. Koopmans, no breve artigo de apresentação *“A Note about Kantorovich’s Paper ‘Mathematical Methods of Production, Planning*

and Organization”⁴⁰, afirma ter sabido dos trabalhos de Kantorovich em 1949.

Em 1959, Kantorovich publicou na Rússia um livro, que foi ampliado em edições posteriores, em que propõe o uso dos métodos da Programação Linear para a planificação da Economia Soviética: *“Economic Calculations of the Best Utilization of Resources”*, resultado de seu curso anual em métodos de otimização e planejamento da Universidade Estatal de Leningrado (GARDNER, p. 640)⁴¹.

8. A programação vetorial: caso linear

Nesta seção estudamos o caso da Programação Linear vetorial. Os problemas básicos da Programação Linear vetorial são:

$$\begin{array}{ll} \max & c_{11} \cdot x_1 + \dots + c_{1n} \cdot x_n \\ & \vdots \\ & c_{p1} \cdot x_1 + \dots + c_{pn} \cdot x_n \\ \text{sa} & a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n \leq b_1; \\ & \vdots \\ & a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n \leq b_m; \\ & x_1 \geq 0, \dots, x_n \geq 0; \end{array}$$

e

$$\begin{array}{ll} \min & c_{11} \cdot x_1 + \dots + c_{1n} \cdot x_n \\ & \vdots \\ & c_{p1} \cdot x_1 + \dots + c_{pn} \cdot x_n \\ \text{sa} & a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n \geq b_1; \\ & \vdots \\ & a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n \geq b_m; \\ & x_1 \geq 0, \dots, x_n \geq 0. \end{array}$$

Observe que são p funções objetivos para serem otimizadas:

$$\begin{array}{l} f_1(x_1, \dots, x_n) = c_{11} \cdot x_1 + \dots + c_{1n} \cdot x_n \\ \vdots \\ f_p(x_1, \dots, x_n) = c_{p1} \cdot x_1 + \dots + c_{pn} \cdot x_n. \end{array}$$

A programação vetorial diferencia-se dos problemas anteriores por uma razão básica: são **várias funções objetivos** $f_1(x), \dots, f_p(x)$ e isso exige que definamos o que se entende por maximizar ou minimizar

este conjunto de funções $f_1(x), \dots, f_p(x)$. A definição empregada é aquela genericamente conhecida pelos economistas como ótimo de Pareto e que vamos chamar de máximos e mínimos vetoriais.

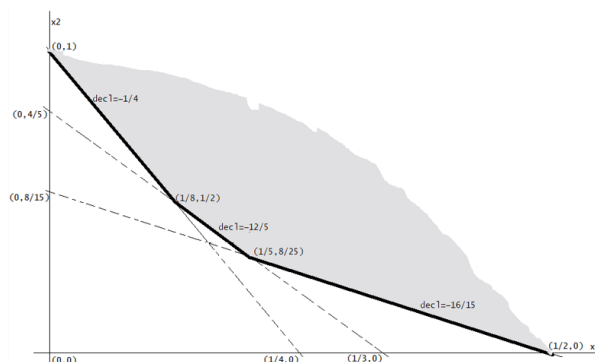
Exemplo 10: Suponhamos que temos suplementos alimentares A e B que possuem três nutrientes importantes para a saúde. Cada 200g de suplemento, tanto de A quanto de B, contêm os nutrientes conforme a tabela abaixo:

Nutriente	A	B	Necessidades mínimas diárias
I	32g	30g	16
II	36g	15g	12
III	64g	16g	6

Digamos que x_1 e x_2 são as unidades de 200g dos suplementos A e B consumidas diariamente, a condição para satisfazer as necessidades diárias é:

$$\begin{aligned} 32 \cdot x_1 + 30 \cdot x_2 &\geq 16 \\ 36 \cdot x_1 + 15 \cdot x_2 &\geq 12 \\ 64 \cdot x_1 + 16 \cdot x_2 &\geq 16 \\ x_1 \geq 0, x_2 &\geq 0. \end{aligned}$$

A região determinada pelas restrições está desenhada abaixo:



Digamos que cada unidade de 200g do alimento A custa R\$ 8,00 e do alimento B custa R\$ 6,50. Nesse caso, se o problema é minimizar

custo e, simultaneamente, satisfazer as restrições, é:

$$\begin{array}{ll} \min & 8 \cdot x_1 + \frac{13}{2} \cdot x_2 \\ \text{sa} & 32 \cdot x_1 + 30 \cdot x_2 \geq 16 \\ & 36 \cdot x_1 + 15 \cdot x_2 \geq 12 \\ & 64 \cdot x_1 + 16 \cdot x_2 \geq 16 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Como a declividade das isoquantas $8 \cdot x_1 + \frac{13}{2} \cdot x_2 = c$ é $\text{decl} = \frac{16}{13}$ e $\frac{16}{15} < \frac{16}{13} < \frac{12}{5}$, basta observar o desenho acima, atentando-se às declividades das retas que estão no gráfico, para ver que a solução está no vértice $(\frac{1}{5}, \frac{8}{25})$, em que o custo é $\text{R\$} \frac{92}{25} = \text{R\$}3,68$, ou seja, deve-se consumir 40g do suplemento A e 64 g do suplemento B diariamente.

Digamos que cada 200g do suplemento A tem 3g de sódio e que cada 200g do suplemento B tem 5g de sódio. O objetivo também é diminuir o consumo de sódio. Os dois objetivos estão em conflito: o alimento mais caro tem menos sódio. O número de gramas de sódio consumido é dado por $3 \cdot x_1 + 5 \cdot x_2$. O problema para minimização do consumo de sódio satisfazendo as necessidades de nutrientes é:

$$\begin{array}{ll} \min & 3 \cdot x_1 + 5 \cdot x_2 \\ \text{sa} & 32 \cdot x_1 + 30 \cdot x_2 \geq 16 \\ & 36 \cdot x_1 + 15 \cdot x_2 \geq 12 \\ & 64 \cdot x_1 + 16 \cdot x_2 \geq 16 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

No caso, como a declividade das isoquantas $3 \cdot x_1 + 5 \cdot x_2 = s$ é $\text{decl} = \frac{3}{5}$ e $\frac{3}{5} < \frac{12}{5}$, então a solução está no vértice $(\frac{1}{2}, 0)$, em que o consumo de 100g do suplemento A implica o menor consumo de sódio que é $\frac{3}{2}$ g. No entanto, no vértice $(\frac{1}{2}, 0)$ o gasto diário é de $\text{R\$}4,00$. Portanto, existe conflito entre os dois objetivos: menor custo, implica maior consumo de sódio.

Se definimos a função vetorial:

$$f(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} 8 \cdot x_1 + \frac{13}{2} \cdot x_2 \\ 3 \cdot x_1 + 5 \cdot x_2 \end{pmatrix}$$

$$\text{ent\~{a}o } f(0,1)=\left(\frac{13}{2}\right)=\left(\frac{6,5}{5}\right), f\left(\frac{1}{8},\frac{1}{2}\right)=\left(\frac{4,25}{2,875}\right), f\left(\frac{1}{5},\frac{8}{25}\right)=\left(\frac{92}{25}\right)=\left(\frac{3,68}{2,2}\right)$$

$$\text{e } f\left(\frac{1}{2},0\right)=\left(\frac{4}{3}\right)=\left(\frac{4}{1,5}\right).$$

Claramente, os v\'ertices $(0,1)$, $\left(\frac{1}{8},\frac{1}{2}\right)$ est\~{a}o eliminados, pois os v\'ertices restantes $\left(\frac{1}{5},\frac{8}{25}\right)$ e $\left(\frac{1}{2},0\right)$ implicam custos menores e menor consumo de s\'odio.

Todavia, dados os dois crit\'erios, n\~{a}o conseguimos decidir entre os v\'ertices $\left(\frac{1}{5},\frac{8}{25}\right)$ e $\left(\frac{1}{2},0\right)$. Os pontos (x'_1,x'_2) , (x_1,x_2) , $\frac{1}{5}\leq x_1,x'_1\leq\frac{1}{2}$, que est\~{a}o sobre a reta $32\cdot x_1+30\cdot x_2=16$, s\~{a}o tais que se $x'_1<x_1$, $x'_2>x_2$, ou seja, os pontos (x'_1,x'_2) e (x_1,x_2) sobre a reta s\~{a}o tais que se (x'_1,x'_2) est\~{a} \~{a} esquerda do ponto (x_1,x_2) , ent\~{a}o $3,68\leq f_1(x'_1,x'_2)<f_1(x_1,x_2)\leq 4$ e $2,2\geq f_2(x'_1,x'_2)>f_2(x_1,x_2)\geq 1,5$. Todos esses pontos apresentam um *tradeoff* entre custo e quantidade de s\'odio e s\~{a}o exemplos de **m\'inimos vetoriais**: n\~{a}o existe a possibilidade de diminuir custo sem aumentar a quantidade de s\'odio e vice-versa.

8.1 Defini\~{c}o de m\'inimos e m\'aximos vetoriais

Seja um conjunto de fun\~{c}oes $f_1,..., f_p:\Omega\subseteq\mathbb{R}^n\rightarrow\mathbb{R}$. Dizemos que $x^*\in\Omega$ \textbf{\'e eficiente no sentido de Pareto} ou **eficiente na defini\~{c}o de Koopmans** ou, ainda, um **m\'inimo vetorial** na defini\~{c}o de Kuhn e Tucker se n\~{a}o existe x vi\'avel, isto \\'e, $x\in\Omega$, tal que

$$\begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \leq \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_p(x^*) \end{pmatrix} \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \neq \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_p(x^*) \end{pmatrix}$$

Ainda, $x^*\in\Omega$ \textbf{\'e um m\'inimo vetorial} se **n\~{a}o existe** $x\in\Omega$ tal que $f_1(x)\leq f_1(x^*)$, $f_2(x)\leq f_2(x^*)$, ..., $f_p(x)\leq f_p(x^*)$ e para algum $1\leq j\leq p$, $f_j(x)<f_j(x^*)$. Em outras palavras, **se n\~{a}o \\'e poss\'ivel diminuir alguns valores $f_j(x)$ em rela\~{c}o \~{a} $f_j(x^*)$ sem que isto implique aumento de outros $f_i(x)$ em rela\~{c}o \~{a} $f_i(x^*)$** . Observe que esse \\'e o caso em rela\~{c}o aos pontos (x_1,x_2) sobre a reta $32\cdot x_1+30\cdot x_2=16$ no segmento de reta entre os pontos $\left(\frac{1}{5},\frac{8}{25}\right)$ e $\left(\frac{1}{2},0\right)$.

De maneira similar, definimos um **máximo vetorial**. Um $x^* \in \Omega$ é dito um máximo vetorial se não existe x viável, isto é, $x \in \Omega$, tal que

$$\begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \geq \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_p(x^*) \end{pmatrix} \text{ e } \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \neq \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_p(x^*) \end{pmatrix}$$

Como anteriormente, $x^* \in \Omega$ é um **máximo vetorial** se **não existe** $x \in \Omega$ tal que $f_1(x) \geq f_1(x^*)$, $f_2(x) \geq f_2(x^*)$, ..., $f_p(x) \geq f_p(x^*)$ e para algum $1 \leq j \leq p$, $f_j(x) > f_j(x^*)$. Em outras palavras, **se não é possível aumentar alguns valores $f_j(x)$ em relação à $f_j(x^*)$ sem que isto implique diminuição de outros $f_i(x)$ em relação à $f_i(x^*)$.**

Genericamente, designamos um mínimo ou máximo vetorial como um **Ótimo de Pareto**.

A definição de Koopmans de ponto eficiente:

um ponto no espaço das *commodities* é dito eficiente se um aumento de uma das suas coordenadas pode ser alcançado apenas pela diminuição de alguma outra. Matematicamente, isto se expressa por [...] um ponto $y \in A$ se não existe um outro $\bar{y} \in A$ tal que $\bar{y} - y \geq 0$ ⁴²

Isto é, $\bar{y}_i \geq y_i$, para todo i e para algum k , $\bar{y}_k > y_k$.

Wilfredo Pareto (1848-1923) detalha o conceito de máximo vetorial no parágrafo 89, “Definição do Máximo de Ofelimidade para uma Coletividade”, do Apêndice do livro “Manual de Economia Política”:

Consideremos uma posição qualquer e suponhamos que dela nos afastemos por uma posição muito pequena de modo compatível com as ligações [restrições]. Se, ao fazê-lo, aumentamos o bem-estar de todos os indivíduos da coletividade, é evidente que a nova posição é mais vantajosa para cada um deles; e vice-versa, ela o é menos se diminuirmos o bem-estar de todos os indivíduos. O bem-estar de alguns deles pode, aliás, permanecer constante, sem que essas conclusões mudem. Mas se, pelo contrário, esse pequeno movimento faz com que aumente o bem-estar de certos indivíduos e diminua de outros, já não podemos afirmar que é vantajoso para toda a coletividade efetuar esse movimento. São essas considerações que levam a definir como posição de máximo de ofelimidade aquela de que é impossível se afastar por uma quantidade muito pequena, sem que todas

as ofelimitudes de que desfrutam os indivíduos, salvo os que permanecem constantes, recebam um aumento ou diminuição.⁴³

Atualmente, a distinção entre “ofelimitude”, que é um neologismo a partir da expressão criada por Pareto, e “utilidade” está eliminada na teoria neoclássica. “Utilidade” substituiu os dois conceitos. Aquilo que hoje designamos como a utilidade associada ao consumidor, por exemplo, ou seja, as utilidades dos agentes econômicos individuais, Pareto chama de “ofelimitudes”.

Conforme a tabela 1 da p. 80 do livro de Tarascio⁴⁴, ofelimitudes podem ser individuais ou coletivas (sociais), mas a fonte de satisfação é necessariamente econômica. Em outras palavras, ofelimitude é uma “medida” de satisfação econômica. Dada a impossibilidade de comparar as ofelimitudes de diferentes indivíduos, somos conduzidos aos chamados ótimos de Pareto, ou seja, às configurações nas quais não podemos aumentar algumas ofelimitudes individuais sem diminuir outras. Isto é, tudo o que é possível graças à inexistência de uma ofelimitude social particular. Apenas os interesses econômicos individuais dos elementos que constituem a coletividade estão disponíveis. Portanto, nossos exemplos e considerações estão nesta classificação: existem interesses econômicos individuais e inexiste uma ofelimitude social junto às individuais. Nestes casos, estamos limitados aos máximos e mínimos vetoriais. É tudo o que é possível do ponto de vista da Economia.

“Utilidade” individual ou social, por sua vez, é uma expressão reservada aos casos em que a fonte de satisfação é qualquer, ou seja, quando a fonte de satisfação não é exclusivamente econômica, podendo se originar em considerações políticas, éticas, estéticas etc. Novamente, Pareto distingue utilidade individual e coletiva, sendo que na utilidade coletiva considera-se apenas as utilidades dos

indivíduos na coletividade sem a existência de uma utilidade própria para a coletividade. Pareto também acresce ao conjunto uma “utilidade social”, que é uma utilidade própria da sociedade sem consideração das utilidades individuais.

O nosso exemplo de minimização vetorial é dado pelo problema:

$$\begin{array}{ll} \min & 8 \cdot x_1 + \frac{13}{2} \cdot x_2 \\ & 3 \cdot x_1 + 5 \cdot x_2 \\ \text{sa} & 32 \cdot x_1 + 30 \cdot x_2 \geq 16 \\ & 36 \cdot x_1 + 15 \cdot x_2 \geq 12 \\ & 64 \cdot x_1 + 16 \cdot x_2 \geq 16 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Em relação a esse exemplo, caso queiramos escolher entre os ótimos definidos pelo segmento de reta $32 \cdot x_1 + 30 \cdot x_2 = 16$ entre os pontos $\left(\frac{1}{5}, \frac{8}{25}\right)$ e $\left(\frac{1}{2}, 0\right)$, temos que estabelecer outro critério de eleição além de custo e quantidade de sódio.

Observemos que a reta $32 \cdot x_1 + 30 \cdot x_2 = \frac{140}{41} \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2\right) + \frac{64}{41} \cdot (3 \cdot x_1 + 5 \cdot x_2)$, ou seja, estamos apresentando o segmento que contém os ótimos como uma combinação de custo e sódio. Digamos que resolvemos testar, dentro do conjunto de ótimos, como diminui a quantidade de sódio quando aumentamos o gasto em R\$1,00:

$$32 \cdot x_1 + 30 \cdot x_2 = \frac{140}{41} \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 + 1\right) + \frac{64}{41} \cdot \left(3 \cdot x_1 + 5 \cdot x_2 - \frac{140}{64}\right).$$

Portanto, um aumento de R\$1,00 no custo implica que o consumo de sódio cai $\frac{140}{64} = 2,1875$ g de sódio, ou seja, cada real gasto a mais aumentando a quantidade consumida do suplemento A e diminuindo o consumo de B implica uma queda de mais de 2g no consumo de sódio. Nessa situação, o consumidor tem a possibilidade de avaliar o seu consumo de sódio e optar por escolher um ponto (x_1, x_2) no segmento que considere adequado.

Um problema de otimização vetorial sempre pode ser reduzido a um problema de otimização escalar: dados $a > 0$ e $b > 0$ quaisquer, a

solução ou soluções do problema

$$\begin{array}{ll} \min & a \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 \right) + b \cdot (3 \cdot x_1 + 5 \cdot x_2) \\ \text{sa} & 32 \cdot x_1 + 30 \cdot x_2 \geq 16 \\ & 36 \cdot x_1 + 15 \cdot x_2 \geq 12 \\ & 64 \cdot x_1 + 16 \cdot x_2 \geq 16 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

É um mínimo vetorial. A declividade das isoquantas $a \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 \right) + b \cdot (3 \cdot x_1 + 5 \cdot x_2) = c$ são dadas pelas frações $-\frac{8 \cdot a + 3 \cdot b}{\frac{13}{2} \cdot a + 5 \cdot b} > -\frac{12}{5}$ para quaisquer $a > 0$ e $b > 0$, em que $-\frac{12}{5}$ é a declividade da reta $36 \cdot x_1 + 15 \cdot x_2 \geq 12$. Portanto, quaisquer que sejam as escolhas $a > 0$ e $b > 0$, então as soluções dos problemas acima são $\left(\frac{1}{5}, \frac{8}{25} \right)$, os pontos (x_1, x_2) sobre $64 \cdot x_1 + 16 \cdot x_2 = 16$, tais que $\frac{1}{5} \leq x_1 \leq \frac{1}{2}$ ou $\left(\frac{1}{2}, 0 \right)$. Por exemplo, se escolhermos $a = \frac{140}{41}$ e $b = \frac{64}{41}$, então $32 \cdot x_1 + 30 \cdot x_2 = \frac{140}{41} \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 \right) + \frac{64}{41} \cdot (3 \cdot x_1 + 5 \cdot x_2)$ e as soluções do problema acima constituem todo o segmento $64 \cdot x_1 + 16 \cdot x_2 = 16$, $\frac{1}{5} \leq x_1 \leq \frac{1}{2}$. Se escolher $a = b = 1$ e $b = \frac{64}{41}$, então $\left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 \right) + (3 \cdot x_1 + 5 \cdot x_2) = 11 \cdot x_1 + \frac{23}{2} \cdot x_2$, cuja declividade é $\text{decl} = -\frac{22}{23} > -\frac{16}{15}$ e a solução está no vértice $\left(\frac{1}{2}, 0 \right)$. Se $a = 3$ e $b = 1$, então $27 \cdot x_1 + \frac{49}{2} \cdot x_2 = 3 \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 \right) + 1 \cdot (3 \cdot x_1 + 5 \cdot x_2)$, cuja declividade é $-\frac{12}{5} < \text{decl} = -\frac{54}{49} < -\frac{16}{15}$ e a solução está no vértice $\left(\frac{1}{5}, \frac{8}{25} \right)$.

Ao escolher $a > 0$ e $b > 0$,

$$a \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 \right) + b \cdot (3 \cdot x_1 + 5 \cdot x_2) = a \cdot \left(8 \cdot x_1 + \frac{13}{2} \cdot x_2 + 1 \right) + b \cdot \left(3 \cdot x_1 + 5 \cdot x_2 - \frac{a}{b} \right),$$

estamos escolhendo um *tradeoff* entre custo e quantidade de sódio: cada real gasto a mais implica uma queda de $\frac{a}{b}$ gramas no consumo de sódio. Por exemplo, se escolher $a = 2$ e $b = 3$, o problema acima é:

$$\begin{array}{ll} \min & 25 \cdot x_1 + 28 \cdot x_2 \\ \text{sa} & 32 \cdot x_1 + 30 \cdot x_2 \geq 16 \\ & 36 \cdot x_1 + 15 \cdot x_2 \geq 12 \\ & 64 \cdot x_1 + 16 \cdot x_2 \geq 16 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

cuja solução está no vértice $\left(\frac{1}{5}, \frac{8}{25} \right)$, pois $\frac{25}{28} < \frac{16}{15}$.

Exemplo 11: Consideremos o problema de maximização vetorial:

$$\begin{array}{ll}
\max & 3 \cdot x_1 + 2 \cdot x_2 \\
& 2 \cdot x_1 + 3 \cdot x_2 \\
sa & 2 \cdot x_1 + x_2 \leq 6 \\
& x_1 + x_2 \leq 4 \\
& x_1 + 4 \cdot x_2 \leq 4 \\
& x_1 \geq 0, x_2 \geq 0.
\end{array}$$

Todos os máximos vetoriais do problema acima são soluções dos problemas

$$\begin{array}{ll}
\max & \theta_1 \cdot (3 \cdot x_1 + 2 \cdot x_2) + \theta_2 \cdot (2 \cdot x_1 + 3 \cdot x_2) \\
sa & 2 \cdot x_1 + x_2 \leq 6 \\
& x_1 + x_2 \leq 4 \\
& x_1 + 4 \cdot x_2 \leq 4 \\
& x_1 \geq 0, x_2 \geq 0,
\end{array}$$

para $\theta_1 > 0$ e $\theta_2 > 0$.

As declividades das restrições são -2, -1 e -1/4. A declividade das isoquantas

$$\theta_1 \cdot (3 \cdot x_1 + 2 \cdot x_2) + \theta_2 \cdot (2 \cdot x_1 + 3 \cdot x_2) = c = (3 \cdot \theta_1 + 2 \cdot \theta_2) \cdot x_1 + (2 \cdot \theta_1 + 3 \cdot \theta_2) \cdot x_2$$

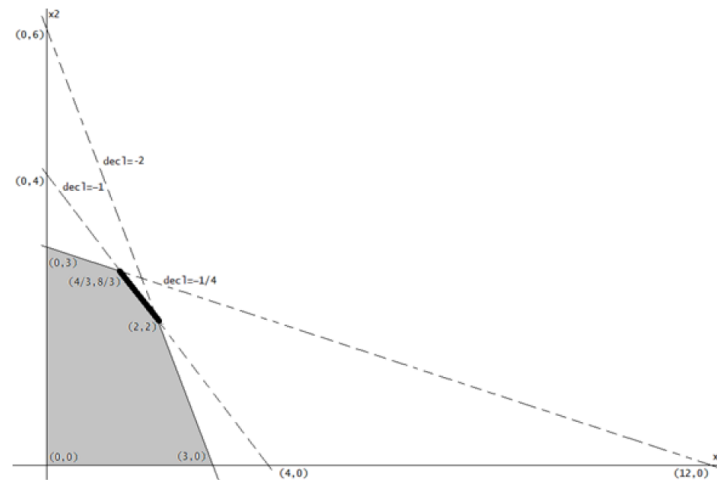
é

$$-\frac{3 \cdot \theta_1 + 2 \cdot \theta_2}{2 \cdot \theta_1 + 3 \cdot \theta_2} = -\frac{3 \cdot \frac{\theta_1}{\theta_2} + 2}{2 \cdot \frac{\theta_1}{\theta_2} + 3}.$$

Observe que $\frac{1}{4} < \frac{3 \cdot \frac{\theta_1}{\theta_2} + 2}{2 \cdot \frac{\theta_1}{\theta_2} + 3} < 2$, portanto, os máximos vetoriais estão no segmento $x_1 + x_2 \leq 4$, $\frac{4}{3} \leq x_1 \leq 2$. Nesse segmento, $\frac{\theta_1}{\theta_2} = 1$. Portanto, ao deslocar-se ao longo desse segmento, uma unidade de $f_1 = 3 \cdot x_1 + 2 \cdot x_2$ está sendo trocada por uma unidade de $f_2 = 2 \cdot x_1 + 3 \cdot x_2$. Suponha que o problema seja um problema de Análise de Atividade e que $f_1 = 3 \cdot x_1 + 2 \cdot x_2$ representa a receita da empresa e que $f_2 = 2 \cdot x_1 + 3 \cdot x_2$ uma medida de preservação ambiental. Portanto, se aumentamos a produção do produto 1, ou seja, x_1 , estamos aumentando a receita $f_1 = 2 \cdot x_1 + 4$ e diminuindo a medida de preservação ambiental que é $f_2 = -x_1 + 12 \cdot x_2$.

Escolhendo $\frac{\theta_1}{\theta_2} > 0$, se está escolhendo quanto à disposição de trocar uma unidade de preservação ambiental por uma unidade de receita.

O resultado será uma escolha de produção no segmento $x_1 + x_2 \leq 4$, $\frac{4}{3} \leq x_1 \leq 2$.



Exercício 15

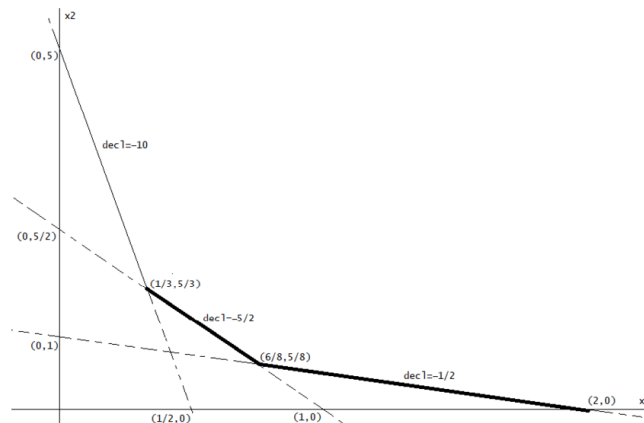
Considere o problema:

$$\begin{array}{ll} \min & x_1 + 3 \cdot x_2 \\ & 3 \cdot x_1 + x_2 \\ \text{sa} & 5 \cdot x_1 + 2 \cdot x_2 \geq 5 \\ & x_1 + 2 \cdot x_2 \geq 2 \\ & 10 \cdot x_1 + x_2 \geq 5 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

semelhante ao exemplo anterior. Podemos determinar todos os mínimos vetoriais resolvendo os problemas:

$$\begin{array}{ll} \min & \theta_1 \cdot (x_1 + 3 \cdot x_2) + \theta_2 \cdot (3 \cdot x_1 + x_2) \\ \text{sa} & 5 \cdot x_1 + 2 \cdot x_2 \geq 5 \\ & x_1 + 2 \cdot x_2 \geq 2 \\ & 10 \cdot x_1 + x_2 \geq 5 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

para todos os $\theta_1 > 0$ e $\theta_2 > 0$. As declividades das retas $\theta_1 \cdot (x_1 + 3 \cdot x_2) + \theta_2 \cdot (3 \cdot x_1 + x_2) = c$ são $-\frac{\theta_1 + 3 \cdot \theta_2}{3 \cdot \theta_1 + \theta_2}$. Observe que $\frac{1}{3} < \frac{\theta_1 + 3 \cdot \theta_2}{3 \cdot \theta_1 + \theta_2} < 3$.



Conclua que os mínimos vetoriais estão sobre os segmentos entre os pontos $(\frac{1}{3}, \frac{5}{3})$, $(\frac{6}{8}, \frac{5}{8})$, $(\frac{6}{8}, \frac{5}{8})$ e $(2, 0)$.

No exemplo anterior, interpretamos $x_1 + 3 \cdot x_2$ como custo e $3 \cdot x_1 + x_2$ como quantidade de sódio. Observe que o suplemento mais barato é o suplemento A. Se utilizamos a mesma interpretação, então $\theta_1 \cdot (x_1 + 3 \cdot x_2) + \theta_2 \cdot (3 \cdot x_1 + x_2) = \theta_1 \cdot (x_1 + 3 \cdot x_2 + 1) + \theta_2 \cdot (3 \cdot x_1 + x_2 - \frac{\theta_1}{\theta_2})$, o que significa que cada real gasto a mais implica uma queda de $\frac{\theta_1}{\theta_2}$ g no consumo de sódio. Se escolhermos algum ponto (x_1, x_2) no segmento $5 \cdot x_1 + 2 \cdot x_2 = 5$ do ponto $(\frac{1}{3}, \frac{5}{3})$ até o ponto $(\frac{6}{8}, \frac{5}{8})$, então $\frac{\frac{\theta_1}{\theta_2} + 3}{3 \cdot \frac{\theta_1}{\theta_2} + 1} = \frac{5}{2}$, portanto, $\frac{\theta_1}{\theta_2} = \frac{1}{13}$.

Nesse caso, cada real a mais no custo representa uma queda de apenas $\frac{1}{13}$ g no consumo de sódio. É possível que o consumidor avalie essa queda no consumo de sódio como insignificante e opte pelo vértice $(\frac{6}{8}, \frac{5}{8})$ no lugar do vértice $(\frac{1}{3}, \frac{5}{3})$, pois $f(\frac{1}{3}, \frac{5}{3}) = (\frac{16}{3}, \frac{8}{3})$ e $f(\frac{6}{8}, \frac{5}{8}) = (\frac{21}{8}, \frac{23}{8})$, a queda de custo é de $\frac{16}{3} - \frac{21}{8} = \frac{65}{24}$, ou seja, quase R\$ 3, enquanto o aumento na quantidade de sódio é de $\frac{23}{8} - \frac{8}{3} = \frac{5}{24}$, ou seja, pouco mais que 0,2g.

Se escolhermos algum ponto (x_1, x_2) no segmento $x_1 + 2 \cdot x_2 = 2$ do ponto $(\frac{6}{8}, \frac{5}{8})$ até o ponto $(2, 0)$, então $\frac{\frac{\theta_1}{\theta_2} + 3}{3 \cdot \frac{\theta_1}{\theta_2} + 1} = \frac{1}{2}$, portanto, $\frac{\theta_1}{\theta_2} = 5$. Nesse caso, cada real a mais no custo representa uma queda de 5g no consumo de sódio. É possível que o consumidor avalie essa queda no consumo de sódio como significativa pois ao passar do vértice $(\frac{6}{8}, \frac{5}{8})$ para o vértice $(2, 0)$, a

queda de custo é de $\frac{21}{8}-2=\frac{5}{8}$, ou seja, menos que R\$ 0,7, enquanto o aumento na quantidade de sódio é $6-\frac{23}{8}=\frac{15}{8}$, ou seja, quase 2g.

Sendo assim, a análise sobre os *tradeoff* entre custo e quantidade de sódio parece indicar que a “melhor” solução é o vértice $(\frac{6}{8}, \frac{5}{8})$. Entretanto, essa é uma das análises possíveis, pois, por exemplo, o consumidor pode optar pelo vértice $(2,0)$ se não tem restrição sobre o consumo de sódio. O que sabemos, com certeza, é que levando em conta dois critérios, custo *versus* quantidade de sódio, o consumidor jamais escolherá algum ponto sobre o segmento $10 \cdot x_1 + x_2 = 5, \quad 0 \leq x_1 < 1/3$, pois todos esses pontos implicam maior custo e maior quantidade de sódio em relação aos demais, que são mínimos vetoriais. Observe que conseguimos $\theta_1 > 0$ e $\theta_2 > 0$, tais que $\frac{\frac{\theta_1}{\theta_2} + 3}{3 \cdot \frac{\theta_1}{\theta_2} + 1} = 10$, pois se $\frac{\frac{\theta_1}{\theta_2} + 3}{3 \cdot \frac{\theta_1}{\theta_2} + 1} = 10$, então $\frac{\theta_1}{\theta_2} = -\frac{7}{29}$. Em outras palavras, se um ponto (x_1, x_2) não é mínimo vetorial, então não é solução de nenhum problema da forma

$$\begin{array}{ll} \min & \theta_1 \cdot (x_1 + 3 \cdot x_2) + \theta_2 \cdot (3 \cdot x_1 + x_2) \\ \text{sa} & 5 \cdot x_1 + 2 \cdot x_2 \geq 5 \\ & x_1 + 2 \cdot x_2 \geq 2 \\ & 10 \cdot x_1 + x_2 \geq 5 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

Com $\theta_1 > 0$ e $\theta_2 > 0$.

10. Comentários históricos sobre otimização vetorial

O resultado que reduz um problema de otimização vetorial a um problema de otimização escalar no contexto da otimização não-linear – através da escolha arbitrária dos $\theta_1 > 0, \dots, \theta_p > 0$ – está no celebrado artigo “*Nonlinear Programming*”, de 1951, de Harold W. Kuhn e Albert W. Tucker⁴⁵. A versão linear é um caso particular do resultado obtido por esses autores.

Em 1972, ocorreu a primeira conferência de **otimização linear vetorial**. O primeiro livro⁴⁶ de Programação Linear multiobjetivo foi

publicado em 1978. Nesse livro, o autor expõe as aplicações da otimização vetorial no uso de recursos hídricos, meio-ambiente, tributação, transporte urbano e relações internacionais.

No capítulo 1, *“Introduction”*, Jared Cohon observa que esses problemas são os mais comuns, principalmente no setor público, quando o planejador trata de vários objetivos sociais representando interesses em conflito – por exemplo, aumento de empregos em determinada região e preservação ambiental, localização de estações de bombeiro em regiões pobres de alta densidade populacional ou regiões industriais com grande concentração de equipamentos caros, porém baixa densidade populacional etc.

A formulação dos problemas em termos de uma função objetivo restringe o papel do tomador de decisões e infla o papel do formulador do modelo. As informações de economistas (quanto ao nível da atividade econômica), engenheiros (quanto aos impactos ambientais, por exemplo), sociólogos (quanto aos impactos no bem-estar da população) são incorporadas aos modelos, fornecendo mais informações e um conjunto amplo de alternativas para o tomador de decisões. A formulação com vários objetivos e soluções via ótimos de Pareto introduz parâmetros que servem para avaliar a sensibilidade das soluções, em termos dos *tradeoffs*, entre as variáveis de escolha. Além disso, os julgamentos de valor implicados nas escolhas são mais explícitos.

No capítulo 9, *“Multiobjective Analysis of Water Resource Problems”*, Cohon trata do problema dos usos da água. Geralmente, a criação de um lago para retenção de água de um ou vários rios considera, do ponto de vista econômico, apenas a relação custo-benefício, ou seja, quanto vai demandar de investimento público, quanto os usuários irão pagar por esses recursos e quanto esses investimentos irão gerar de renda – que via os impostos amortizam os investimentos públicos

iniciais. Entretanto, esse critério é demasiadamente restritivo. Por exemplo, os investimentos iniciais são menores em regiões economicamente mais desenvolvidas por causa da infraestrutura existente nessas regiões. Provavelmente exista um interesse público em desenvolver regiões economicamente menos desenvolvidas. Além disso, quando a infraestrutura é construída numa região determinada, implica custo para outras. Um lago diminui o fluxo de água em certas regiões ou expõe-nas a inundações quando as comportas precisam liberar água. Como Cohon alerta, mesmo atividades complementares, por exemplo, um lago para geração de energia hidroelétrica e irrigação de lavouras, podem entrar em conflito: a necessidade de abrir comportas pode inundar regiões antes da colheita ou, em épocas de estiagem, a represa pode reter a água que a agricultura necessita. Somado a todos esses problemas, temos os impactos ambientais que podem afetar negativamente regiões distantes do projeto.

Mesmo na captação de água tanto para uso doméstico quanto industrial existem conflitos, pois a água que retorna para o rio ou para o reservatório depois do uso industrial pode estar comprometida pra uso humano.

Exercício 16

Dado o problema de minimização vetorial

$$\begin{array}{ll} \min & 2 \cdot x + y \\ & x + 2 \cdot y \\ \\ \text{sa} & 3 \cdot x + y \geq 6 \\ & x + y \geq 4 \\ & x + 3 \cdot y \geq 6 \\ & x \geq 0, y \geq 0, \end{array}$$

Determine todos os ótimos de Pareto do problema e a fração $\frac{\theta_1}{\theta_2}$ de maneira a obter todos estes ótimos.

Exercício 17

Dado o problema de minimização vetorial

$$\begin{array}{ll}\min & x + y \\ & x + 6 \cdot y \\ \\ & 3 \cdot x + y \geq 6 \\ sa & x + y \geq 4 \\ & x + 3 \cdot y \geq 6 \\ & x \geq 0, y \geq 0,\end{array}$$

Determine todos os ótimos de Pareto do problema e a fração $\frac{\theta_1}{\theta_2}$ de maneira a obter todos estes ótimos.

Exercício 18

Considere o problema

$$\begin{array}{ll}\min & x_1 + 2 \cdot x_2 + 3 \cdot x_3 \\ & x_1 + x_2 \\ & 3 \cdot x_1 + 2 \cdot x_2 + x_3 \\ sa & x_1 + x_2 + 3 \cdot x_3 \geq 3 \\ & 4 \cdot x_1 + 4 \cdot x_2 + 6 \cdot x_3 \geq 9 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0.\end{array}$$

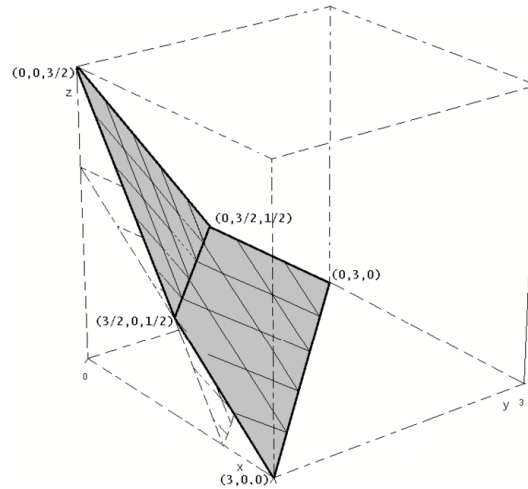
Os máximos vetoriais do problema acima são soluções dos problemas escalares abaixo para $\theta_1 > 0$, $\theta_2 > 0$, $\theta_3 > 0$:

$$\begin{array}{ll}\min & \theta_1 \cdot (x_1 + 2 \cdot x_2 + 3 \cdot x_3) + \theta_2 \cdot (x_1 + x_2) + \theta_3 \cdot (3 \cdot x_1 + 2 \cdot x_2 + x_3) \\ sa & x_1 + x_2 + 3 \cdot x_3 \geq 3 \\ & 4 \cdot x_1 + 4 \cdot x_2 + 6 \cdot x_3 \geq 9 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0\end{array}$$

ou

$$\begin{array}{ll}\min & (\theta_1 + \theta_2 + 3 \cdot \theta_3) \cdot x_1 + (2 \cdot \theta_1 + \theta_2 + 2 \cdot \theta_3) \cdot x_2 + (3 \cdot \theta_1 + \theta_3) \cdot x_3 \\ sa & x_1 + x_2 + 3 \cdot x_3 \geq 3 \\ & 4 \cdot x_1 + 4 \cdot x_2 + 6 \cdot x_3 \geq 9 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0.\end{array}$$

i. Verifique que os vértices da região viável são o que aparecem no desenho abaixo:



ii,. Faça $\theta_1=8$, $\theta_2=1$ e $\theta_3=3$. Verifique que os pontos (x_1, x_2, x_3) do segmento de reta $x = \frac{3}{2} - \frac{3}{2} \cdot t$, $y=0$, $z = \frac{1}{2} + t$ para $0 \leq t \leq 1$ são mínimos vetoriais.

Observe que fixados $\theta_1 > 0$, $\theta_2 > 0$, $\theta_3 > 0$, se

$F(x_1, x_2, x_3) = (\theta_1 + \theta_2 + 3 \cdot \theta_3) \cdot x_1 + (2 \cdot \theta_1 + \theta_2 + 2 \cdot \theta_3) \cdot x_2 + (3 \cdot \theta_1 + \theta_3) \cdot x_3$, então

$$F\left(\frac{3}{2}, 0, \frac{1}{2}\right) = 3 \cdot \theta_1 + \frac{3 \cdot \theta_2}{2} + 5 \cdot \theta_3$$

$$F\left(0, 0, \frac{3}{2}\right) = \frac{9 \cdot \theta_1}{2} + \frac{3 \cdot \theta_3}{2}$$

$$F\left(0, \frac{3}{2}, \frac{1}{2}\right) = \frac{9 \cdot \theta_1}{2} + \frac{3 \cdot \theta_2}{2} + \frac{7 \cdot \theta_3}{2}$$

$$F(3, 0, 0) = 3 \cdot \theta_1 + 3 \cdot \theta_2 + 9 \cdot \theta_3$$

$$F(0, 3, 0) = 6 \cdot \theta_1 + 3 \cdot \theta_2 + 6 \cdot \theta_3.$$

A partir dos valores $F(x_1, x_2, x_3)$ conclua que $\left(0, \frac{3}{2}, \frac{1}{2}\right)$, $(3, 0, 0)$ e $(0, 3, 0)$ não são ótimos de Pareto.

iv. Conclua que os ótimos que encontramos no item ii. são todos máximos vetoriais desse problema.

¹ COTTLE, R. "A Brief History of the International Symposia on Mathematical Programming". **Mathematical Programming, Serie B**, nº 125 (2010), pp. 207-233

² DÜPPE, T.; WEINTRAUB, R. "Siting the New Economic Science: The Cowles Commission's Activity Analysis Conference of June 1949". **European Historical Economies Society (EHES) Working Paper**, nº 40 (2013).

- DÜPPE, T.; WEINTRAUB, R. **Finding Equilibrium, Arrow, Debreu, McKenzie and the Problem of Scientific Credit**. New York: Princeton University Press, 2014, p. 98.
- ³ KOOPMANS, T. "Introduction". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, p. 1.
- ⁴ DÜPPE, T.; WEINTRAUB, R. "Siting the New Economic Science: The Cowles Commission's Activity Analysis Conference of June 1949". **European Historical Economics Society (EHES) Working Paper**, nº 40 (2013), p. 4.
- ⁵ KANTOROVICH, L. "Mathematical Methods of Production Planning and Organization". In: NEMCHINOV, V. S. **The Use of Mathematics in Economics**. Londres: Oliver & Boyd, 1964, pp. 225-243
- ⁶ HITCHCOCK, F. L. "The Distribution of a Product from Several Sources to Numerous Localities". **Journal of Mathematics and Physics**, nº 20 (1941), pp. 224-230.
- ⁷ KOOPMANS, T. "Optimum Utilization of the Transportation System". **Econometrica**, v. 17 (1949), suplemento, pp. 136-146.
- ⁸ KOOPMANS, T. "Introduction". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, p. 4.
- ⁹ DORFMAN, R.; SAMUELSON, P.; SOLOW, R. **Linear Programming and Economic Analysis**. New York: Dover Publications, INC., 1987, p. 106.
- ¹⁰ STIGLER, G. "The Cost of Subsistence". **Journal of Farm Economics**, v. 27, nº 2 (1945), pp. 303-314.
- ¹¹ DANTZIG, G. "Reminiscences about the Origins of Linear Programming". **Operations Research Letters**, v. 1, nº 2 (1982), p. 47.
- ¹² DANTZIG, G. "Reminiscences about the Origins of Linear Programming". **Operations Research Letters**, v. 1, nº 2 (1982), p. 45.
- ¹³ KOOPMANS, T. "Introduction". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, p. 330.
- ¹⁴ COTTLE, R. "George Dantzig: A Legendary Life in Mathematical Programming". **Mathematical Programming**, v. 105, nº 1 (2006), p. 6.
- ¹⁵ COTTLE, R. "Operations Research Icon". **Operations Research**, v. 53 (2005), nº 6, p. 896
- ¹⁶ KOOPMANS, T. "Analysis of Production as an Efficient Combination of Activities in Activity Analysis of Production and Allocation" [Capítulo III]. In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, p. 33.
- ¹⁷ Ibidem, pp. 34-35.
- ¹⁸ Ibidem, p. 1.
- ¹⁹ WALD, A. "On the Unique Non-Negative Solvability of the New Production Equations (Part 1), On the Production Equations of Economic Value Theory (Part 2)". In: BAUMOL,

- W.; GOLDFELD, S. **Precursors in Mathematical Economics**: An Anthology. Londres: The London School and Political Science, 1968, pp. 278-293.
- WALD, A. "On Some Systems of Equations of Mathematical Economics" **Econometrica**, v. 19, nº 4 (1951), pp. 360-403 [publicada inicialmente em 1936].
- ²⁰ NEUMANN, J. "A Model of General Economic Equilibrium". **The Review of Economic Studies**, v. 13, nº 1 (1945), pp. 1-9.
- ²¹ LEONTIEF, W. "Quantitative Input and Output Relations in the Economic Systems of the United States". **The Review of Economics and Statistics**, v. 18, nº 3 (1936), pp. 105-125.
- ²² KOOPMANS, T. "Introduction". In: KOOPMANS, T. **Activity Analysis of Production and Allocation, Proceedings of a Conference**. New York: John Wiley & Sons, INC., 1951, p. 4.
- ²³ LOUÇÃ, F. **The Years of High Econometrics**: A Short History of the generation that Reinvented Economics. New York: Routledge, 2007, pp. XXV-VI.
- ²⁴ WERIN, L.; JUNGELFELT, K. "Tjalling Koopman's Contribution to Economics". **The Scandinavian Journal of Economics**, v. 78, nº 1 (1976), p. 90.
- ²⁵ GUGLIELMO, M. "The Contribution of Economists to Military Intelligence During World War II". **The Journal of Economic History**, v. 68, nº 1 (2008), p. 144.
- ²⁶ CHRIST, Carl. "The Cowles Commission's Contributions to Econometrics at Chicago". **Journal of Economic Literature**, v. 32, nº 1 (1994), p. 32.
- ²⁷ Ibidem, p. 31.
- ²⁸ TOBIN, J. "Portrait: Tjalling Koopmans". **Challenge**, v. 20, nº 4 (1977), p. 59.
- ²⁹ STIGLER, G. "The Cost of Subsistence". **Journal of Farm Economics**, v. 27, nº 2 (1945), pp. 303-314.
- ³⁰ DORFMAN, R.; SAMUELSON, P.; SOLOW, R. **Linear Programming and Economic Analysis**. New York: Dover Publications, Inc., 1987, p. 9.
- ³¹ DANTZIG, G. "Reminiscences about the Origins of Linear Programming". **Operations Research Letters**, v. 1, nº 2 (1982) p. 45.
- ³² Hitchcock, F. L. "The Distribution of a Product from Several Sources to Numerous Localities". **Journal of Mathematics and Physics**, nº 20 (1941), pp. 224-230.
- ³³ Ibidem, p. 227.
- ³⁴ DANTZIG, G. "The Programming of Interdependent Activities: Mathematical Model". In: KOOPMANS, T. **Activity Analysis of Production and Allocation, Proceedings of a Conference**. Cowles Commission Monograph No. 13. NY: John Wiley & Sons., 1951, p. 31.
- ³⁵ Ibidem, p. 4.
- ³⁶ KANTOROVICH, L. "Mathematical Methods of Production Planning and Organization". In: NEMCHINOV, V. S. **The Use of Mathematics in Economics**. Londres: Oliver & Boyd, 1964, pp. 230.
- ³⁷ GARDNER, R. "L. V. Kantorovich: The Price Implications of Optimal Planning". **Journal of Economic Literature**, v. 28, nº 2 (1990).

- ³⁸ KANTOROVICH, L. "Mathematical Methods of Production Planning and Organization". In: NEMCHINOV, V. S. **The Use of Mathematics in Economics**. Londres: Oliver & Boyd, 1964, pp. 226.
- ³⁹ GARDNER, R. "L. V. Kantorovich: The Price Implications of Optimal Planning". *Journal of Economic Literature*, v. 28, nº 2 (1990), pp. 643.
- ⁴⁰ KOOPMANS, T. "A Note about Kantorovich's Paper 'Mathematical Methods of Production, Planning and Organization'". **Management Science**, v. 6, nº 4 (1960), pp. 363-365.
- ⁴¹ GARDNER, R. "L. V. Kantorovich: The Price Implications of Optimal Planning". **Journal of Economic Literature**, v. 28, nº 2 (1990), pp. 640.
- ⁴² KOOPMANS, T. "Analysis of Production as an Efficient Combination of Activities". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, p. 60.
- ⁴³ PARETO, W. **Manual de Economia Política**. V. II [Capítulos VI-IX e Apêndice]. SP: Abril Cultural, Coleção Os Economistas, 1984, p. 159.
- ⁴⁴ TARASCIO, V. **Pareto's Methodological Approach to Economics**. Chapel Hill: The University of North Carolina Press, 1968.
- ⁴⁵ KUHN, H.; TUCKER, A. "Nonlinear Programming". In: NEWMAN, P. (Ed.). **Readings in Mathematical Economics**. V. I. Baltimore: The John Hopkins Press, 1968, p. 3.
- ⁴⁶ COHON, J. **Multiobjective Programming and Planning**. Londres: Academic Press, INC, 1978.

Lição DOIS

A Teoria dos Jogos de Soma-Zero

Sumário

Lição 2 – Teoria dos Jogos de Soma-Zero

1. Introdução
2. Breve histórico
3. Jogos de soma-zero
 - 3.1 Quando as soluções não são vértices: escolha qualquer um!
 - 3.2 Determinando selas
 - 3.4 A Teoria dos Jogos vai à guerra (Exemplo 14)
 - 3.5 A interpretação da solução como um conjunto de probabilidades
4. A polêmica
5. O guia von Neumann para Economia

Lição 2

Teoria dos Jogos de Soma-Zero

1. Introdução

O objetivo desta lição é apresentar jogos de soma-zero tais como foram definidos por John von Neumann no famoso artigo de 1928, assim como os equilíbrios definidos como pares de pontos definidos nos simplexes básicos Δ_{m-1} e Δ_{n-1} e definindo selas para os *payoffs* esperados pelos jogadores.

A seguir, expomos esses jogos como um par de problemas, primal e dual, de programação linear, tal como foi apresentado por George Dantzig¹. Encerramos a lição com alguns comentários sobre a existência de múltiplas selas e definimos estratégias dominantes, quando existem, para certas situações nas quais algumas selas não parecem opções razoáveis para os jogadores.

O conceito de sela como pares de estratégias ótimas foi posteriormente generalizado por John Forbes Nash (1928-2015) e hoje recebe o nome de “Equilíbrio de Nash”. Então, o conceito de selas que veremos a seguir são exemplos de Equilíbrios de Nash para o caso particular de jogos de soma-zero com dois jogadores.

Do ponto de vista da teoria que estamos analisando, os jogos de soma-zero com dois jogadores tiveram um papel essencial. Foram esses jogos que evidenciaram a importância da dualidade na programação linear. O artigo “*Nonlinear Programming*”, de 1951, de

Kuhn e Tucker, que é nosso objetivo central, foi desenvolvido sob a forma de um teorema da dualidade para funções não lineares.

O emprego do conceito de conjunto convexo utilizado por von Neumann no artigo "*A Model of General Economic Equilibrium*", de 1932, e no livro "*The Theory of Games*", de 1944, foi responsável pelo desenvolvimento da Análise Convexa para a Teoria da Otimização e para a Teoria Econômica.

2. Breve histórico

O marco inicial da Teoria dos Jogos é um artigo analisando o jogo de xadrez², do matemático alemão Ernest Zermelo, publicado em 1913. O trabalho de Zermelo explora os aspectos formais do jogo. Ele se pergunta sobre o número mínimo de movimentos necessários para vencer uma partida se o jogador está numa posição que lhe permita vencer.

Na década de 1920, Émile Borel publicou três importantes artigos nos quais alerta para o fato de que certas circunstâncias na guerra ou na economia podem ser vistas como jogos. Foi ele quem primeiramente definiu o que hoje chamamos equilíbrio em estratégias mistas e soluções maxmin e minimax³.

Émile Borel (1871-1956) foi um jovem precoce que publicou seus primeiros trabalhos matemáticos com apenas 19 anos. Em 1893, com 22 anos, concluiu o doutorado na área com a tese "*Sur quelques Points de la Théorie des Fonctions*", que foi um trabalho importante para a Matemática do século 20.

Durante a 1ª Guerra, Borel comandou um batalhão de artilharia. Seu único filho, que também foi combatente, faleceu no conflito.

Em 1928, Borel fundou o Instituto Poincaré, que hoje abriga o centro Émile Borel, com a ajuda da Fundação Rockefeller e dos

banqueiros Rothschild. Entre 1905 e 1950, Borel publicou em torno de 50 trabalhos em teoria das probabilidades. Borel preocupou-se principalmente com as aplicações em Física, Matemática, Biologia e Ciências Sociais.

A partir de 1924, a atividade política de Borel tornou-se mais intensa, sendo eleito deputado em 1928. Borel foi ministro da marinha da França entre 1925 e 1940. Mesmo durante esse período, o trabalho matemático de Borel não foi interrompido e continuou também durante a 2ª Guerra.

Em 1941, foi preso pelo regime de Vichy por causa das suas atividades na resistência francesa. Em 1946, aos 75 anos, publicou um livro famoso, “Os Paradoxos do Infinito”, e continuou publicando nos anos seguintes. Em 1948, foi eleito presidente do Comitê de Ciências da Unesco.

Entretanto, Borel não conseguiu definir equilíbrios para jogos em geral, tarefa que foi completada por John von Neumann no famoso artigo publicado em 1928⁴.

John von Neumann (1903-1957) foi outro prodígio matemático que também publicou trabalhos ainda adolescente. Outro matemático húngaro, Paul Halmos⁵ (1916-2006), que foi assistente de von Neumann, escreveu um relato repleto de pequenas histórias sobre a prodigiosa memória e a capacidade de fazer cálculos de cabeça de von Neumann. Aos sete anos, já sabia cálculo. Segundo Halmos, também húngaro, o matemático Peter Lax (1926) – mais jovem entre os colaboradores de von Neumann –, comentou que certa vez alguém perguntou a von Neumann quanto uma pessoa podia aspirar saber de matemática no século 20. Depois de pensar alguns instantes, von Neumann respondeu: em torno de 28%!

Von Neumann foi um campeão no uso do método matemático do século 20, que é o chamado “método axiomático”, cujas origens

remontam ao final do século 19 e começo do século 20, com o matemático alemão David Hilbert (1862-1943). Sob orientação de Hilbert, von Neumann permaneceu durante um ano em pós-doutorado, entre 1926 e 1927, na Universidade de Göttingen. O artigo sobre jogos foi uma tarefa apresentada nos seminários orientados por Hilbert. Os problemas que atraíam von Neumann no começo da sua carreira pertenciam à teoria dos conjuntos, à lógica matemática e à mecânica quântica – problemas que ocuparam Hilbert e sua escola.

Von Neumann trabalhou no Projeto Manhattan e, em 1955, foi diagnosticado com câncer ósseo, causado pela exposição à radiação ao assistir explosões atômicas no Pacífico, segundo alguns relatos ⁶. Ele morreu no dia 8 de fevereiro de 1957.

Em 1944, von Neumann, em coautoria com o economista austríaco Oskar Morgenstern, publicou aquele que é um dos monumentos intelectuais do século 20: “A Teoria dos Jogos e do Comportamento Econômico”⁷. Esse livro revolucionou a teoria econômica tanto sob o aspecto metodológico, quanto nos novos conceitos e instrumentos formais que disponibilizou para a teoria econômica. Essa obra é o marco inicial do uso da Teoria dos Jogos na teoria econômica e, de maneira geral, nas ciências.

Os primeiros jogos estudados eram não-cooperativos, no sentido de que os objetivos dos jogadores são conflitantes, como nos jogos de soma-zero. Esse é o caso de jogos como o xadrez ou o pôquer. Contudo, podemos imaginar jogos em que existe um incentivo à cooperação entre os participantes. A teoria destes jogos também deriva tanto do artigo de 1928, quanto do livro de von Neumann e Morgenstern, que modelam esses jogos a partir da Teoria dos Jogos não-cooperativos.

Von Neumann e Morgenstern não nutriam simpatia em relação à Teoria Neoclássica e à Teoria do Equilíbrio Geral, e suas críticas estão expostas no capítulo I de *“Theory of Games and Economic Behavior”*. Além disso, os autores projetaram a Teoria dos Jogos como uma alternativa para a Teoria Neoclássica. Em 1973, para uma enciclopédia, Morgenstern resumiu sua crítica e suas intenções no verbete *“Game Theory”*:

A justificativa essencial para tomar os jogos de estratégia como modelos para grandes classes de comportamento humano já foi estabelecida [...]: nossos atos são interdependentes de maneiras muito complexas e é a forma precisa dessa interdependência que deve ser estabelecida. A interdependência, é claro, foi reconhecida, mas mesmo onde a economia neoclássica do tipo Walras-Pareto tentou descrever essa interdependência, a tentativa falhou porque não havia nenhum método rigoroso para explicar a interação, o que é evidente especialmente quando o número de agentes é pequeno, como no oligopólio (poucos vendedores). Em vez disso, introduziu-se um grande número de participantes (sob o nome impróprio de “livre competição”) de forma que, assintoticamente, nenhum tem qualquer influência perceptível sobre qualquer outro participante e, conseqüentemente, sobre o resultado, cada um enfrentando apenas condições fixas. Assim, a suposta tarefa do indivíduo era apenas maximizar seu lucro ou utilidade, em vez de considerar as atividades dos “outros”. Em vez de resolver o problema econômico dado empiricamente, ele foi negado; mas a realidade não desaparece. Na política internacional, é evidente que nunca há mais do que alguns Estados, nos parlamentos alguns partidos, nas operações militares alguns exércitos, divisões, navios, etc. Portanto, as unidades de decisão eficazes tendem a permanecer pequenas. A interação das decisões permanece mais óbvia e falta uma teoria rigorosa.⁸

Entre os anos de 1950 e 1953, o matemático americano John Nash (1928-2015), então um jovem com pouco mais de 21 anos, publicou quatro célebres artigos que quatro décadas depois o premiariam com o Nobel de Economia⁹. Nesses artigos, Nash generalizou o conceito de equilíbrio em maxmin-minmax, de Borel e Newmann, para jogos não-cooperativos sem a restrição de soma constante e propôs o estudo de jogos cooperativos. Todos esses artigos são desdobramentos de sua tese de doutorado *“Non-Cooperative Games”*, defendida no Departamento de Matemática de Princeton em maio

de 1950. O orientador de Nash foi Albert Willian Tucker (1905-1995), um dos responsáveis por nosso assunto: o Teorema de Kuhn-Tucker.

O primeiro livro-texto de Teoria dos Jogos, de autoria do matemático John Charles McKinsey (1908-1953), é publicado em 1952, com o título de *“Introduction to the Theory of Games”*¹⁰.

Em 1982, o biólogo John Maynard Smith (1920-2004) publicou o livro *“Evolution and the Theory of Games”*, que inicia a aplicação da Teoria dos Jogos à Biologia¹¹.

Em 1990, o economista David Kreps (1950) publicou o livro *“A Course in Microeconomic Theory”*, que é o primeiro livro de Microeconomia a introduzir a Teoria dos Jogos num texto voltado para pós-graduação¹².

Em 1994, Douglas Baird, Robert Gertner e Randal Picker publicaram *“Game Theory and the Law”*, que é o primeiro livro a aplicar a Teoria dos Jogos ao Direito¹³.

Em 1994, John Nash (1928-2015), John Harsanyi (1920-2000) e Reinhard Selten (1930-2016) receberam o Nobel de Economia pelas suas contribuições à Teoria dos Jogos.

3. Jogos de soma-zero

Na seção de 14 de maio de 1928, Émile Borel (1871-1956) informou à Academia de Ciências a respeito de uma nota dirigida à instituição pelo jovem matemático John von Neumann (NEUMANN, 1928). Na nota, von Neumann comunicava que resolvera um problema proposto por Borel em relação a uma questão sobre jogos: “Qual é a melhor maneira de jogar tanto para o jogador I quanto para o jogador II?”. Na nota, Neumann afirmou que resolvera o problema numa forma generalizada e criticou Borel:

No seu livro “Éléments du Calcul des Probabilités” como também em várias notas publicadas nos “*les Comptes Rendus*” [Relatórios], M. E. Borel ocupou-se do seguinte problema: “Dois jogadores, I e II, participam de um jogo que consiste em I escolher um número x e II, um número y ; a escolha de x é feita entre os números $1, 2, \dots, M$ e a escolha de y , entre os números $1, 2, \dots, N$. Cada um dos jogadores escolhe sem saber qual a escolha do outro; se I e II escolheram, respectivamente, x e y , II deve pagar a_{xy} ($a_{xy} \geq 0, a_{xy} \leq 0$) para I; a matriz $\{a_{xy}\}$, ($x=1, 2, \dots, M; y=1, 2, \dots, N$) caracteriza o jogo: é a regra do jogo. Quais são as melhores maneiras de jogar tanto para I quanto para II? [pergunta-se Borel] [...] Tendo me dedicado independentemente a este problema (um pouco generalizado), obtive, entre outras coisas, uma resposta afirmativa para a questão principal (não resolvida) colocada por M. Borel [...]. A dedução detalhada dos resultados serão publicados logo num outro periódico: *Zur Theorie der Gesellschaftsspiele*. [...] Uma observação essencial é a seguinte: a pergunta “como deve o jogador I (resp. II) jogar, para garantir o maior ganho possível, ou seja, obter o maior a_{xy} (resp., menor) possível?” não tem [para Borel] um significado definido. De fato, como pode I fazer (resp., II) a_{xy} grande (resp., pequeno) visto que I não tem o poder de fixar o valor [de a_{xy}] – I determina apenas x (resp., y), o índice da linha (resp., coluna) da matriz $\{a_{xy}\}$? [...] O verdadeiro problema é encontrar uma definição exata e plausível (dos interesses de I e II), permitindo uma solução única. ¹⁴

Entre 1921 e 1927, Borel¹⁵ publicou três artigos famosos em Teoria dos Jogos em que define estratégias como um conjunto de probabilidades e obtém, em casos particulares, os equilíbrios como selas dos *payoffs* associados aos jogadores. Todavia, Borel não conseguiu definir equilíbrios para jogos em geral. Essa última tarefa foi realizada por John von Neumann¹⁶ ao adotar os conceitos de estratégias mistas e de estratégias maxmin e minmax introduzidas por Borel no artigo de 1921 (BOREL, 1953a). Von Neumann admitiu que Borel formulara a hipótese maxmin e minmax, mas que a verificara apenas em alguns casos particulares. Por sua vez, von Neumann provava a validade dessa hipótese para todos os jogos de dois jogadores sem a necessidade de se restringir aos jogos chamados simétricos definidos por Borel.

Consideremos um jogo de soma-zero para dois jogadores dado por uma matriz

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Por convenção, um dos jogadores, dito o jogador das linhas, escolhe uma linha da matriz; o outro, dito o jogador das colunas, escolhe uma coluna. Escolhida uma linha, digamos i , por parte do jogador das linhas, e uma coluna, digamos j , por parte do jogador das colunas, o *payoff* do jogador das linhas é a_{ij} e o do jogador das colunas é $-a_{ij}$. Portanto, a matriz A é a matriz dos *payoffs* do jogador das linhas; $-A$ é a matriz dos *payoffs* do jogador das colunas; chamamos de jogos de soma-zero porque a soma dos *payoffs* é zero.

Exemplo 1: Seja $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 2 \end{pmatrix}$. Se o jogador das linhas escolher a linha dois e o jogador das colunas, a coluna dois, o jogador da linha recebe -1, o que significa que ele está pagando +1 real para o jogador das colunas. Por outro lado, se o jogador das linhas escolher a linha três e o jogador das colunas, a coluna um, significa que o jogador das linhas está recebendo +1 real do jogador das colunas.

O problema é determinar como esses jogadores procedem em suas escolhas, já que, obviamente, o objetivo dos dois jogadores é maximizar seus ganhos ou – o que é a mesma coisa – minimizar suas perdas.

A solução proposta por von Neumann, seguindo Borel, é a seguinte: o jogador das linhas observa em cada linha qual é o pior cenário. Na linha um, o pior cenário é o jogador das colunas escolher a coluna dois. Nesse caso, o *payoff* do jogador das linhas é zero. Feito isso para cada linha, resultam-se os *payoffs* $\{0, -1, 1\}$, que é o conjunto dos piores resultados nas linhas. Dado que esses são os piores resultados para o jogador das linhas, ele escolhe a linha três, que é o cenário “menos ruim”.

Mais tecnicamente, aquilo a que chamamos o pior cenário em cada linha é o que definimos como o nível de segurança desta linha, ou seja, o jogador das linhas sabe com certeza que não receberá um valor menor que aquele ao escolher essa linha. Na linha um, o jogador das linhas não terá que pagar nada na pior das hipóteses. Na linha 2, o jogador das linhas não gastará mais que um real na pior das hipóteses. Portanto, no nosso exemplo, os níveis de segurança são $\{0, -1, 1\}$. O jogador das linhas escolhe a linha três, cujo nível de segurança é 1, garantindo assim um *payoff* de no mínimo 1. Digamos que o nível de segurança da linha i é $v_i = \min_j a_{ij}$, isto é, jogando a linha i , o jogador das linhas recebe no mínimo v_i , então o jogador maximiza o seu nível de segurança escolhendo a linha onde está o maior v_i , isto é, $\max_i v_i = \max_i \min_j a_{ij}$. Essa é a chamada **estratégia max-min** para o jogador das linhas.

Por sua vez, o jogador das colunas procede da mesma maneira; seus níveis de segurança são $\{1, 2\}$. Portanto, ele escolhe a coluna um, pois, jogando essa coluna, assegura com certeza que não pagará mais que 1. Digamos que o nível de segurança da coluna j é $u_j = \max_i a_{ij}$, isto é, jogando a linha j o jogador das colunas não paga mais que o máximo u_j , então o jogador minimiza o seu nível de segurança escolhendo a linha onde está o menor u_j , isto é, $\min_j u_j = \min_j \max_i a_{ij}$. Essa é a chamada **estratégia min-max** para o jogador das colunas.

Caso, os níveis de segurança dos dois jogadores se igualem, isto é, $\max_i v_i = \min_j u_j = v$, então existem escolhas de linha e coluna que asseguram para ambos o melhor nível de segurança comum v , que é dito o valor do jogo. Essas são as chamadas soluções maxmin-minmax para jogos de soma-zero. Costuma-se abreviar por “**solução maxmin**”.

Resultado no exemplo 1: o jogador das linhas escolhe a linha três, o jogador das colunas escolhe a coluna um e o jogador das linhas recebe +1 do jogador das colunas. A solução do jogo, ou seja, a linha

escolhida, a coluna escolhida e o valor do jogo são obtidos respeitando a igualdade

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = v.$$

Exemplo 2: Seja o jogo $A = \begin{pmatrix} 3 & 5 & -5 & 0 \\ 7 & 5 & 9 & \boxed{2} \\ 3 & 2 & 1 & -5 \\ 2 & -8 & 3 & 2 \end{pmatrix}$. Os níveis de segurança para o jogador das linhas são: $\{-5, 2, -5, -8\}$, $\max\{-5, 2, -5, -8\} = 2$. O jogador das linhas escolhe a linha dois. Observe que em todas as outras linhas, o nível de segurança é menor, existindo o risco de ter de pagar para o jogador das colunas. Os níveis de segurança para o jogador das colunas são $\{7, 5, 9, 2\}$, $\min\{7, 5, 9, 2\} = 2$. O jogador das colunas escolhe a coluna quatro. Portanto, o jogador das linhas escolhe a linha dois e o jogador das colunas escolhe a coluna quatro e paga R\$ 2 para o jogador das linhas.

Exemplo 3: Consideremos o jogo $\Gamma = \begin{pmatrix} 1 & -1 & -\frac{1}{2} \\ 2 & -2 & 0 \\ 3 & 5 & -3 \end{pmatrix}$. A partir de um jogo Γ , von Neumann define dois outros jogos: Γ_1 , que chama de jogo minorante associado ao jogo Γ , e Γ_2 , que designa por majorante associado ao Γ , na seção 14.2¹⁷.

Γ_1 é definido da seguinte maneira: o jogador das colunas escolhe após o jogador das linhas, sabendo qual é a escolha do jogador das linhas – que por sua vez, escolhe sabendo que o jogador das colunas tem ciência da sua escolha. Nesse caso, sem dúvida, o jogador das linhas escolhe a estratégia maxmin: se escolher a linha um, dois ou três perderá -1 , -2 ou -3 , respectivamente. A menor das perdas é $\max \min = -1$. Uma vez que o jogador das linhas escolheu, o jogador das colunas seleciona a coluna dois e impõe uma perda de 1 real ao jogador das linhas. Portanto, a solução do jogo Γ_1 é $\max_j \min_i a_{ij}$. Isto é, o jogador das linhas procura nas linhas as “piores” colunas e depois a coluna

“menos ruim”, que impõe ao jogador das colunas. Esse é o resultado $\max_j \min_i a_{ij}$. O jogador das linhas, escolhendo antes, impõe ao jogador das colunas o resultado que lhe é menos desfavorável.

Γ_2 é definido de forma similar: o jogador das linhas escolhe após o jogador das colunas, sabendo qual é a escolha do jogador das colunas – que por sua vez, escolhe sabendo que o jogador das linhas tem ciência da sua escolha. Nesse caso, sem dúvida, o jogador das colunas escolhe a estratégia minmax: se escolher a coluna um, dois ou três perderá 3, 5 ou 0, respectivamente. Obviamente, selecionando $\min_{\max} = 0$ não paga nada. Uma vez que o jogador das colunas escolheu, o jogador das linhas seleciona a linha dois e nenhum deles têm perda. Portanto, a solução do jogo Γ_2 é $\max_i \min_j a_{ij}$. Isto é, o jogador das colunas procura nas colunas as “piores” linhas e depois a linha “menos ruim”, que será a escolha do jogador das linhas. Esse é o resultado $\min_i \max_j a_{ij}$. Na verdade, o jogador das colunas, escolhendo antes, impõe ao jogador das linhas o resultado que lhe é menos desfavorável.

Observe! Se o jogador das colunas tiver conhecimento sobre as escolhas do jogador das linhas, ou seja, se estiverem jogando Γ_1 , o jogador das linhas “aceita” perder R\$ 1. Se, por outro lado, se estiverem jogando Γ_2 , o jogador das linhas sabe como o jogador das colunas irá selecionar, ou seja, se o jogador das colunas tiver que jogar antes, para impedir perdas maiores, escolhe a coluna três e aceita não ganhar nada, pois sabe que o jogador das linhas escolherá a linha dois. Claro que se o jogador das linhas tiver conhecimento sobre a maneira como atua o jogador das colunas, então o jogador das linhas evita a perda de R\$ 1 e não paga nada ao jogador das colunas. Portanto, uma conclusão importante que tiramos dos jogos Γ_1 e Γ_2 é que

$$\max_j \min_i a_{ij} \leq \min_i \max_j a_{ij},$$

pois, obviamente, se o jogador das colunas souber da escolha do jogador das linhas, este está em desvantagem; por outro lado, se o jogador das linhas souber com antecedência qual a seleção do jogador das colunas, ele estará na posição vantajosa. Caso o jogo Γ seja efetivado, não sabemos qual será o resultado, ou seja, quanto alguém paga ou quanto recebe. Efetivamente, ao término do jogo, o jogador das linhas recebe algum valor Γ_{ij} que corresponde a algum elemento (i, j) da matriz Γ . Não é possível determinar Γ_{ij} a priori. Por sua vez, o jogador das colunas recebe o valor $-\Gamma_{ij}$. Podemos conjecturar que nesses casos existe um “resultado esperado” v que deve estar entre os dois casos extremos de total desvantagem para um ou outro jogador: $\max_j \min_i a_{ij} \leq v \leq \min_i \max_j a_{ij}$. Nesse exemplo, $v = -0,6$, para sabermos obter o valor v , temos que aguardar a seção 3.2.

Exemplo 4: Voltemos ao jogo $A = \begin{pmatrix} 3 & 5 & -5 & 0 \\ 7 & 5 & 9 & \boxed{2} \\ 3 & 2 & 1 & -5 \\ 2 & -8 & 3 & 2 \end{pmatrix}$. Os níveis de segurança para o jogador das linhas são $\{-5, 2, -5, -8\}$, $\max\{-5, 2, -5, -8\} = 2$. Seguindo a notação anterior, o resultado do jogo minorante A_1 é 2; o resultado do jogo majorante A_2 também é 2. Portanto, nesse jogo, nenhum dos dois jogadores aufere vantagem por saber com antecedência a escolha do oponente. Von Neumann chama os jogos como o jogo A de **estritamente determinados**. Nesses casos, é como se o valor esperado v ficasse “espremido” entre $\max_j \min_i a_{ij} \leq v \leq \min_i \max_j a_{ij}$.

Nos jogos tipo Γ , como no exemplo 3, o segredo é importante. Se o jogador das colunas souber da opção do jogador das linhas, como em Γ_1 , ele eleva seus ganhos, diminuindo suas perdas. Portanto, nossos jogos estão divididos em duas classes: os jogos em que o segredo não desempenha nenhum papel em que os lances podem ser

sequenciais e aqueles em que o segredo é importante, em que os lances sequenciais resultam em resultados diversos.

Exercício 1

i) Considere o jogo $A = \begin{pmatrix} 4 & 5 & 4 \\ 3 & 0 & 1 \end{pmatrix}$. Utilizando o método maxmin-minmax $(\max_i \min_j a_{ij} = \min_j \max_i a_{ij})$ acima, conclua que $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ são estratégias ótimas para os jogadores das linhas e das colunas; conclua que $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ também são estratégias ótimas para os jogadores. Portanto, esse jogo tem no mínimo duas soluções.

ii) Utilizando o método maxmin-minmax $(\max_i \min_j a_{ij} = \min_j \max_i a_{ij})$ acima, conclua que $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ são estratégias ótimas para os jogadores das linhas e das colunas, respectivamente, para o jogo $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \end{pmatrix}$; o valor do jogo é $v=1$.

iii) Utilizando o método maxmin-minmax para o jogo $A = \begin{pmatrix} 4 & 4 & 1 \\ 0 & -2 & 3 \\ 6 & 5 & 7 \end{pmatrix}$, conclua que as estratégias ótimas para os jogador das linhas e colunas são $r = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ e $c = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, respectivamente.

iv) Utilizando o método maxmin-minmax para o jogo $A = \begin{pmatrix} 4 & 5 & 4 \\ 5 & 0 & 1 \end{pmatrix}$, conclua que as estratégias ótimas para os jogadores das linhas e colunas são $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, respectivamente.

Exercício 2

Considere o jogo chamado “*matching pennies*” $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. O jogador das linhas tem um único nível de segurança que é -1 e o jogador das colunas, o nível de segurança é 1 então

$$-1 = \max_r \min_c r' \cdot A \cdot c < \min_c \max_r r' \cdot A \cdot c = 1.$$

O que se pode dizer sobre o jogo minorante A_1 e sobre o jogo majorante A_2 ? O segredo é importante nesse jogo? Esse jogo não tem solução da maneira como definimos até aqui. Responda a mesma questão em relação ao jogo chamado “rock-scissors-paper” $A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$.

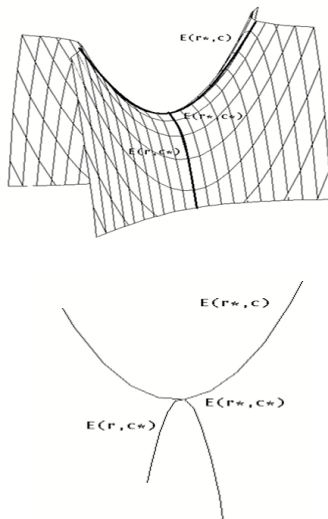
Seja o jogo $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$. Digamos que o jogador das linhas escolha a linha i , isto é, $r_i = e_i = \begin{pmatrix} \vdots \\ 1 \end{pmatrix} \leftarrow i$; é o vértice i do simplexo Δ_{m-1} , ou seja, aquele vetor que tem 1 (um) na posição i e 0 (zero) nas demais posições. Digamos que c é a escolha de coluna, por parte do jogador das colunas, seja a coluna j , ou seja, $c_j = e_j = \begin{pmatrix} \vdots \\ 1 \end{pmatrix} \leftarrow j$, isto é, um vértice do simplexo $\Delta_{n-1} \subseteq \mathbb{R}^{12}$, então temos $m \times n$ números $E_{ij} = E(r_i, c_j) = a_{ij} = r_i' \cdot A \cdot c_j = \left(\cdots \underset{i}{1} \cdots \right) \cdot \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} \vdots \\ 1 \end{pmatrix} \leftarrow j$. Vamos designar genericamente um número desses como $E(r, c) = r' \cdot A \cdot c$, dadas as escolhas de vértices $r \in \Delta_{m-1}$ e $c \in \Delta_{n-1}$. Observe que o *payoff* do jogador das linhas sempre é um número a_{ij} destes e que o *payoff* do jogador das colunas é $-a_{ij}$.

Digamos que determinado jogo seja estritamente determinado. Imaginemos que o jogador das colunas escolhe c^* , então o seu nível de segurança em c^* é $u(c^*) = \max_r r' \cdot A \cdot c^* \geq r' \cdot A \cdot c^*$, isto é, o jogador das colunas não paga mais que $\max_r r' \cdot A \cdot c^*$ dada a sua escolha c^* . Se, numa coluna ou linha, definíssemos um nível de segurança, este seria o pior resultado que o correspondente jogador pode naquela coluna ou linha. Voltando, $v = r'' \cdot A \cdot c^* = u(c^*) = \max_r r' \cdot A \cdot c^* \geq r' \cdot A \cdot c^*$, para qualquer r , portanto, $r'' \cdot A \cdot c^* \geq r' \cdot A \cdot c^*$.

Por outro lado, se o jogador das linhas escolhe r^* então o seu nível de segurança em r^* é $v(r^*) = \min_c r^* \cdot A \cdot c$, isto é, o jogador das linhas não recebe menos que $\min_c r^* \cdot A \cdot c$, ainda, $v = r'' \cdot A \cdot c^* = v(r^*) = \min_c r'' \cdot A \cdot c \leq r'' \cdot A \cdot c$, para qualquer c , portanto, $r'' \cdot A \cdot c^* \leq r'' \cdot A \cdot c$. Reunindo os dois resultados, obtemos:

$$\boxed{r' \cdot A \cdot c^* \leq r'' \cdot A \cdot c^* \leq r'' \cdot A \cdot c.}$$

Mais adiante, comentaremos detalhadamente este resultado que caracteriza a **solução maxmin-minmax** do jogo (r^*, c^*) como um **ponto de sela da função** $E(x, y) = x' \cdot A \cdot y$, ou simplesmente, sela.



Observe que $r' \cdot A \cdot c^* \leq r' \cdot A \cdot c \leq r'' \cdot A \cdot c$ está dizendo que o elemento a_{ij} da matriz A , que é o *payoff* do jogador das linhas no caso de jogos estritamente determinados, é tal que na linha i em que se encontra é o menor elemento, $r' \cdot A \cdot c^* \leq r' \cdot A \cdot c$, e na coluna j em que se encontra é o maior elemento, $r' \cdot A \cdot c^* \leq r'' \cdot A \cdot c^*$.

Exercício 3

Observe que o jogo $A = \begin{pmatrix} 8 & 5 & 6 \\ 0 & -2 & 3 \\ 7 & 5 & 8 \end{pmatrix}$ tem duas selas. Soluções maxmin-minmax não são únicas.

Inversamente, se (r^*, c^*) é uma **sela de** $E(x, y) = x' \cdot A \cdot y$, $\min_c r' \cdot A \cdot c \leq r' \cdot A \cdot c^* \leq r'' \cdot A \cdot c^*$, então $\max_r \min_c r' \cdot A \cdot c \leq r'' \cdot A \cdot c^*$; por outro lado, $\max_r r' \cdot A \cdot c \geq r'' \cdot A \cdot c \geq r'' \cdot A \cdot c^*$, logo $\min_c \max_r r' \cdot A \cdot c \geq r'' \cdot A \cdot c^*$. Juntando os dois resultados:

$$\max_r \min_c r' \cdot A \cdot c \leq r'' \cdot A \cdot c^* \leq \min_c \max_r r' \cdot A \cdot c. (*)$$

Por outro lado, como $r'' \cdot A \cdot c^* \leq r'' \cdot A \cdot c$, então $r'' \cdot A \cdot c^* \leq \min_c r'' \cdot A \cdot c \leq \max_r \min_c r' \cdot A \cdot c$; como $r' \cdot A \cdot c^* \leq r'' \cdot A \cdot c^*$, então $\max_r r' \cdot A \cdot c^* \leq r'' \cdot A \cdot c^*$ e, portanto, $\min_c \max_r r' \cdot A \cdot c \leq \max_r r' \cdot A \cdot c^* \leq r'' \cdot A \cdot c^*$. Conclui-se que:

$$\min_c \max_r r' \cdot A \cdot c \leq r'' \cdot A \cdot c^* \leq \max_r \min_c r' \cdot A \cdot c. (**)$$

Portanto, através de (*) e (**), conclui-se que $r'' \cdot A \cdot c^* = \max_r \min_c r' \cdot A \cdot c = \min_c \max_r r' \cdot A \cdot c$. Logo, os jogadores têm à disposição nessa sela pelo menos uma solução max-min e min-max.

Consideremos dois exemplos muito usados por von Neumann.

Exemplo 5: Considere o jogo chamado “*matching pennies*” $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. O jogador das linhas tem um único nível de segurança, que é -1, e para o jogador das colunas, o nível de segurança é 1, então:

$$v_1 = -1 = \max_r \min_c r' \cdot A \cdot c < \min_c \max_r r' \cdot A \cdot c = 1 = v_2.$$

Portanto, esse jogo não tem selas que sejam estratégias puras, isto é, (1,0) ou (0,1). Esse jogo não tem solução da maneira como definimos até aqui.

Exemplo 6: Considere o jogo chamado “*rock-scissors-paper*” $A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$. O jogador das linhas tem um único nível de segurança, que é -1, e para o jogador das colunas, o nível de segurança é 1, então existem exemplos de jogos em que

$$v_1 = -1 = \max_r \min_c r' \cdot A \cdot c < \min_c \max_r r' \cdot A \cdot c = 1 = v_2.$$

Portanto, esse jogo, como o anterior, não tem selas. Esse problema será resolvido ampliando o conjunto de estratégias para os jogadores das linhas e das colunas, de maneira que passem a existir estratégias que sejam selas para os *payoffs* esperados.

Exemplo 7: Considere o jogo $A = \begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & 1 \\ 1 & -2 & 0 \end{pmatrix}$. A estratégia max-min do jogador das linhas é $r = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, mas esse jogo não tem uma sela, pois $v_1 = -1 = \max_r \min_c r' \cdot A \cdot c < \min_c \max_r r' \cdot A \cdot c = 1 = v_2$. Se o jogador das linhas insistir em sua estratégia max-min, está correndo risco de ter que pagar R\$ 1 para o

jogador das colunas. Se o jogador das colunas souber que o jogador da linha jogará max-min (o jogo minorante), então escolherá a coluna um. Nesse jogo, o segredo é importante. Por outro lado, se o jogador das colunas jogar antes do jogador das linhas e optar pela escolha mais segura (o jogo majorante), a escolha recairá sobre a coluna um e pagará provavelmente R\$ 1 para o jogador das linhas. A teoria que temos, até este ponto, não nos permite determinar estratégias para ambos os jogadores.

Exemplo 8: Como observamos acima, soluções maxmin-minmax não são únicas. Seja $A = \begin{pmatrix} 4 & 5 & 4 \\ 3 & 0 & 1 \end{pmatrix}$. Os níveis de segurança para o jogador das linhas são 4 e 0, portanto, esse jogador escolhe a linha um para garantir um *payoff* de no mínimo 4. Por sua vez, o jogador das colunas, cujos níveis de segurança são 4, 5 e 4, pode escolher indiferentemente a coluna um ou a coluna três. Portanto, nesse exemplo, a menos que o jogador das colunas tenha uma informação adicional que lhe permita discriminar entre as colunas um e três, as duas escolhas garantem o mesmo nível de segurança, que é 4. Se o jogador das colunas quer se proteger de um provável desvio da estratégia max-min por parte do jogador das linhas, então, obviamente, deve jogar a coluna três. Mas, o jogador das colunas sabe, com quase certeza, que escolhendo a coluna um ou a coluna dois pagará R\$ 4 para o jogador das linhas, pois não é razoável que o jogador das linhas se desvie da estratégia max-min. Há apenas uma certeza: jamais o jogador das colunas escolherá coluna dois. Esse exemplo está no livro “*Games and Decisions, Introduction and Critical Survey*” (1957), de R. Duncan Luce e Howard Raiffa¹⁸.

Exercício 4

Consideremos o jogo $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 2 \end{pmatrix}$. As soluções obtidas estão nos vértices $r^* = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ e $c^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, então $r^{*T} \cdot A \cdot c^* = 1$. Observe que $(r_1 \ r_2 \ r_3) \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = r_1 + r_3 \leq 1$ e $1 \leq c_1 + 2 \cdot c_2 = (0 \ 0 \ 1) \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$. Portanto, (r^*, c^*) é uma sela para E , em que $E(r, c) = (r_1 \ r_2 \ r_3) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$.

3.1 Quando as soluções não são vértices: escolha qualquer um!

Em relação à aparente inexistência de soluções para alguns exemplos anteriores, existem duas possibilidades. Abandona-se a definição de soluções maxmin-minmax ou altera-se o conjunto de soluções como foi proposto até o momento. Tanto Borel quanto von Neumann optaram por ampliar os conjuntos de estratégias disponíveis para os jogadores.

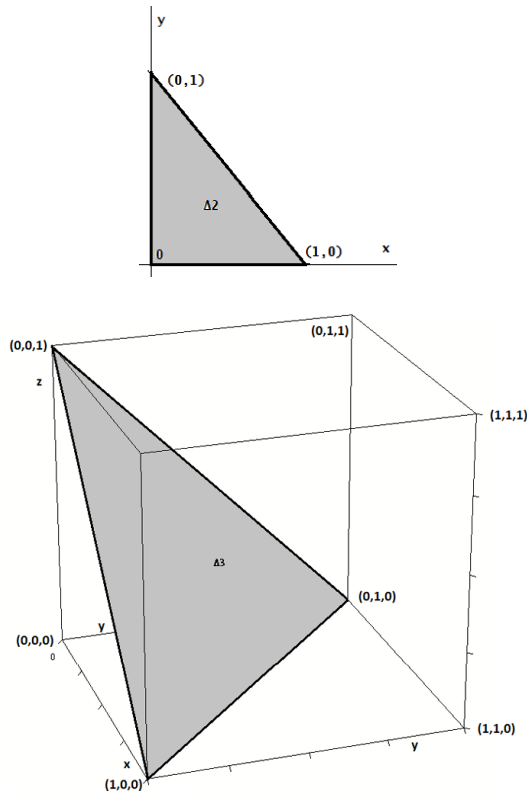
Na comunicação de 1928¹⁹, von Neumann apresenta dois exemplos que já apresentamos anteriormente: $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ e $A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$. Em relação às matrizes acima, temos: $\max_i \min_j a_{ij} = -1 < 1 = \min_j \max_i a_{ij}$. Von Neumann observa que sempre vale $\max_i \min_j a_{ij} \leq \min_j \max_i a_{ij}$, mas que nem sempre vale a igualdade, como nos exemplos acima.

Para remediar essa situação, von Neumann propõe a solução elaborada por Borel: cada jogador escolhe um conjunto $(p_1 \ p_2 \ \dots \ p_k)$, em que $p_i \geq 0$, $p_1 + p_2 + \dots + p_k = 1$ e k é o número de linhas ou colunas à disposição do jogador. Esses conjuntos podem ser representados pelos simplexes

$\Delta_k = \left\{ \begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} \in \mathbb{R}^k \mid r_1 + \dots + r_k = 1, r_i \geq 0 \right\}$, cuja interpretação trataremos mais adiante. Abaixo

estão os simplexes básicos e $\Delta_2 = \left\{ \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in \mathbb{R}^2 \mid c_i \geq 0, c_1 + c_2 = 1 \right\}$ $\Delta_3 = \left\{ \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \in \mathbb{R}^3 \mid r_i \geq 0, r_1 + r_2 + r_3 = 1 \right\}$,

representados pelos seguintes triângulos:



Exemplo 9: No exemplo $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 2 \end{pmatrix}$, a solução que encontramos atribui para o jogador das linhas o vértice $(0,0,1)$ do simplexo Δ_3 e para o jogador das colunas, o vértice $(1,0)$ do simplexo Δ_2 . Da mesma maneira, no jogo $A = \begin{pmatrix} 2 & 0 \\ 2 & 1 \\ 2 & 3 \end{pmatrix}$ é evidente que o jogador das linhas escolhe o vértice $(0,0,1)$ e jogador das colunas o vértice $(1,0)$.

Entretanto, como estamos vendo, não é verdade que para qualquer jogo de soma-zero sempre teremos soluções maxmin e minmax definidas nos vértices dos respectivos simplexos. Isso apenas é possível para jogos estritamente determinados.

Exemplo 10: Se o jogo é dado pela matriz $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$, observe que os procedimentos maxmin e minmax empregados anteriormente não definem nenhum vértice nem para o jogador das linhas, nem para o jogador das colunas. Em cada linha, por exemplo, o menor valor é 1

e em cada coluna o maior valor é 3. Portanto, pelo procedimento que definimos anteriormente, não existe como definirmos uma solução. Qual linha escolhe o jogador das linhas ou qual coluna, o jogador das colunas?

Exemplo 11: A diferença entre o jogo $A = \begin{pmatrix} 2 & 0 \\ 2 & 1 \\ 2 & 3 \end{pmatrix}$ e o jogo $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$ é que no primeiro jogo, supondo que ambos os jogadores que sejam prudentes, ou seja, não se expõem a riscos desnecessários, cada jogador sabe como jogará o oponente, mas não obtém vantagem desse conhecimento. Entretanto, no segundo jogo, se um jogador souber com antecedência como jogará o oponente, fará sua própria escolha de forma a prejudicá-lo e obter maior ganho.

A ideia genial de von Neumann, e antes dele de Émile Borel, é não abandonar os procedimentos maxmin e minmax e aprimorar a ideia de “escolha qualquer vértice”. Mesmo na escolha aleatória das linhas ou colunas, temos como discriminar entre os vértices.

Novamente, consideremos os simplexes básicos $\Delta_3 = \left\{ \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \in \mathbb{R}^3 \mid r_i \geq 0, r_1 + r_2 + r_3 = 1 \right\}$ e $\Delta_2 = \left\{ \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in \mathbb{R}^2 \mid c_i \geq 0, c_1 + c_2 = 1 \right\}$. Definimos a função abaixo que chamamos de *payoff* esperado:

$$E(r, c) = r' \cdot A \cdot c = (r_1 \ r_2 \ r_3) \cdot \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = (3 \cdot r_1 + r_2 + r_3) \cdot c_1 + (r_1 + 2 \cdot r_2 + 3 \cdot r_3) \cdot c_2 > 0.$$

Supomos que o jogador das linhas tem por objetivo maximizar $E(r, c)$ e que, ao contrário, o jogador das colunas busque minimizar $E(r, c)$.

Suponhamos que $r^* = (r_1^*, r_2^*, r_3^*) \in \Delta_3$ e $c^* = (c_1^*, c_2^*) \in \Delta_2$ são as estratégias ótimas para o jogador das linhas e para o jogador das colunas, respectivamente. Ora, isso implica que não é possível que o jogador das linhas obtenha resultado melhor com outra estratégia $r = (r_1, r_2, r_3)$ do que aquele resultado que obtém com $r^* = (r_1^*, r_2^*, r_3^*)$, caso o jogador das colunas escolha

$c^* = (c_1^*, c_2^*)$, isto é, $E(r, c^*) \leq E(r^*, c^*)$. Chamamos $v^* = E(r^*, c^*)$ o valor do jogo. Suponhamos que $v^* > 0$. Portanto,

$$E(e_1, c^*) = 3 \cdot c_1^* + c_2^* \leq v^*$$

$$E(e_2, c^*) = c_1^* + 2 \cdot c_2^* \leq v^*$$

$$E(e_3, c^*) = c_1^* + 3 \cdot c_2^* \leq v^*.$$

Lembrando que $e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i$.

Se dividirmos as expressões anteriores por $v^* > 0$, obtemos

$$3 \cdot \left(\frac{c_1^*}{v^*} \right) + \left(\frac{c_2^*}{v^*} \right) \leq 1$$

$$\left(\frac{c_1^*}{v^*} \right) + 2 \cdot \left(\frac{c_2^*}{v^*} \right) \leq 1$$

$$\left(\frac{c_1^*}{v^*} \right) + 3 \cdot \left(\frac{c_2^*}{v^*} \right) \leq 1.$$

Sejam $x_1 = \frac{c_1^*}{v^*}$ e $x_2 = \frac{c_2^*}{v^*}$. Observe que $x_1 + x_2 = \frac{c_1^* + c_2^*}{v^*} = \frac{1}{v^*}$. O objetivo do jogador das colunas é obter o menor valor possível para v^* . Portanto, o objetivo do jogador das colunas é maximizar $\frac{1}{v^*}$, ou ainda, maximizar $x_1 + x_2$. As restrições acima se expressam como

$$3 \cdot x_1 + x_2 = 3 \cdot \left(\frac{c_1^*}{v^*} \right) + \left(\frac{c_2^*}{v^*} \right) \leq 1$$

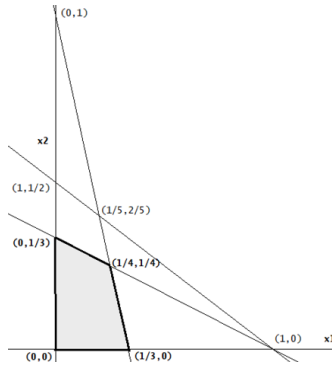
$$x_1 + 2 \cdot x_2 = \left(\frac{c_1^*}{v^*} \right) + 2 \cdot \left(\frac{c_2^*}{v^*} \right) \leq 1$$

$$x_1 + 3 \cdot x_2 = \left(\frac{c_1^*}{v^*} \right) + 3 \cdot \left(\frac{c_2^*}{v^*} \right) \leq 1.$$

Em resumo, o jogador das colunas precisa resolver o problema:

$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{sa} & 3 \cdot x_1 + x_2 \leq 1 \\ & x_1 + 2 \cdot x_2 \leq 1 \\ & x_1 + 3 \cdot x_2 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

cuja solução está em $\left(\frac{1}{4}, \frac{1}{4} \right)$ (verifique isso!), $x_1 + x_2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = \frac{1}{v^*}$, logo $c_1^* = x_1 \cdot v^* = \frac{1}{4} \cdot 2 = \frac{1}{2} = x_2 \cdot v^* = c_2^*$. Portanto, $c^* = \left(\frac{1}{2}, \frac{1}{2} \right)$.



Por outro lado, caso o jogador das linhas escolha $r^* = (r_1^*, r_2^*, r_3^*)$, não é possível para o jogador das colunas obter um resultado melhor com outra estratégia $c = (c_1, c_2)$ que aquele resultado obtido com $c^* = (c_1^*, c_2^*)$, ou seja, $v^* = E(r^*, c^*) \leq E(r^*, c)$. Portanto,

$$E(r^*, e_1) = 3 \cdot r_1^* + r_2^* + r_3^* \geq v^*$$

$$E(r^*, e_2) = r_1^* + 2 \cdot r_2^* + 3 \cdot r_3^* \geq v^*.$$

Dividindo as expressões anteriores por $v^* > 0$, obtemos

$$3 \cdot \left(\frac{r_1^*}{v^*} \right) + \left(\frac{r_2^*}{v^*} \right) + \left(\frac{r_3^*}{v^*} \right) \geq 1$$

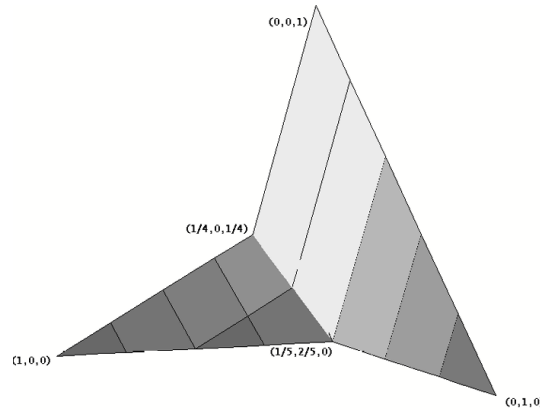
$$\left(\frac{r_1^*}{v^*} \right) + 2 \cdot \left(\frac{r_2^*}{v^*} \right) + 3 \cdot \left(\frac{r_3^*}{v^*} \right) \geq 1.$$

Sejam $y_1 = \frac{r_1^*}{v^*}$, $y_2 = \frac{r_2^*}{v^*}$ e $y_3 = \frac{r_3^*}{v^*}$. Observe que $y_1 + y_2 + y_3 = \frac{r_1^* + r_2^* + r_3^*}{v^*} = \frac{1}{v^*}$. O objetivo do jogador das linhas é obter o maior valor possível para v^* . Portanto, seu objetivo é minimizar $\frac{1}{v^*}$, ou ainda, resolver o problema:

$$\begin{array}{ll} \min & y_1 + y_2 + y_3 \\ \text{sa} & 3 \cdot y_1 + y_2 + y_3 \geq 1 \\ & y_1 + 2 \cdot y_2 + 3 \cdot y_3 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

cuja solução está em $(\frac{1}{4}, 0, \frac{1}{4})$ (verifique isso!), $y_1 + y_2 + y_3 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = \frac{1}{v^*}$, logo

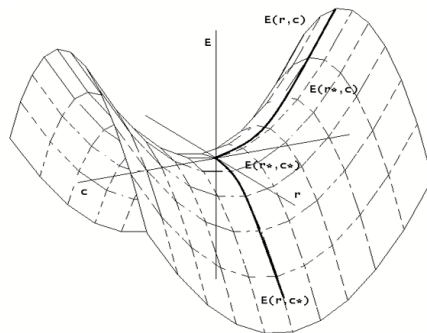
$$r_1^* = y_1 \cdot v^* = \frac{1}{4} \cdot 2 = \frac{1}{2} = y_3 \cdot v^* = r_3^* \text{ e } r_2^* = y_2 \cdot v^* = 0. \text{ Portanto, } r^* = \begin{pmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix}.$$



Observe que as “soluções” $r^* \in \Delta_3$ e $c^* \in \Delta_2$ são tais que

$$E(r, c^*) \leq E(r^*, c^*) \leq E(r^*, c).$$

Um par (r^*, c^*) que satisfaz a condição acima é dito uma **sela** da função E . Como já observamos anteriormente:



As “soluções” que buscamos para os jogos são selas da função E . O procedimento que empregamos é o seguinte: fixado $r \in \Delta_3$, buscamos o $c(r) \in \Delta_2$ que minimiza $E(r, c)$, observe que $c \in \Delta_2$ depende do $r \in \Delta_3$ fixado, por isso escrevemos $c(r)$ – isso corresponde à etapa “min” do maxmin. Por fim, procuramos o $r^* \in \Delta_3$ que maximiza $E(r, c(r))$, ou seja, a etapa “max”. Um procedimento análogo empregamos para obter $c^* \in \Delta_2$: fixado $c \in \Delta_2$, buscamos o $r(c) \in \Delta_3$ que maximiza $E(r, c)$, observe que $r \in \Delta_3$ depende do $c \in \Delta_2$ fixado, por isso escrevemos $r(c)$ – isso corresponde à etapa max do minmax. Por fim, procuramos o $c^* \in \Delta_2$ que minimiza $E(r(c), c)$.

Portanto, von Neumann, ao propor soluções através das chamadas estratégias mistas, estende para além dos vértices dos triângulos, chamados de estratégias puras, a mesma formulação das soluções obtida anteriormente, ou seja, as selas dos *payoffs* esperados. A diferença está em que alguns jogos apresentam soluções nos vértices e outros não. A necessidade dessa extensão é ditada pela estrutura matemática do jogo. Essa estrutura determina que em certos jogos não é necessário esconder do oponente como jogar e em outros jogos o segredo é essencial. Num jogo cujas estratégias ótimas são estratégias puras, o segredo não é importante. Em jogos cujas soluções envolvem estratégias mistas, o segredo é essencial para se preservar de perdas. Para definir “segredo”, precisamos atribuir um significado às estratégias mistas e isso não fizemos ainda. Von Neumann observa no parágrafo 17.1.1 do *“Theory of Games and Economic Behavior”*:

Anteriormente, destacamos que, por exemplo, no jogo do “Matching Pennies”, nenhuma maneira particular de jogar, isto é, nem “cara” e nem “coroa” é melhor que a outra opção e tudo o que importa é descobrir as intenções do oponente. Isso parece bloquear o caminho para uma solução, uma vez que as regras do jogo em questão proíbem explicitamente cada jogador do conhecimento sobre as ações do oponente, no momento em que ele faz sua escolha. Mas a observação acima não corresponde totalmente à realidade do caso: **Ao jogar “Matching Pennies” contra um oponente moderadamente inteligente, o jogador não tentará descobrir as intenções do oponente, mas se concentrará em evitar que o oponente descubra as suas, jogando irregularmente “cara” e “coroa” em jogos sucessivos. Como desejamos descrever a estratégia em uma jogada – de fato, devemos discutir o andamento de uma jogada e não o de uma sequência de jogadas sucessivas – é preferível expressar o seguinte: a estratégia do jogador não consiste em jogar “coroa” ou “cara”, mas em jogar “coroa” e “cara” com probabilidade $\frac{1}{2}$.**²⁰ (grifo e tradução do autor).

Para von Neumann, as estratégias seguras necessariamente estão situadas em selas do *payoff* esperado e uma “maneira particular” de jogar é se um jogador ou os dois jogadores dispõem de estratégias puras em selas. Se, por outro lado, ambos os jogadores não dispõem

de estratégias puras em selas, então precisam esconder do oponente sua escolha. Se não existe uma estratégia pura à disposição, o segredo é essencial. Uma técnica para garantir o segredo é que a opção seja feita aleatoriamente. No parágrafo 17.2.1 do livro “*Theory of Games and Economic Behavior*”, Von Neumann e Morgenstern continuam:

Nesta configuração [ou seja, simplexos de probabilidades e *payoff* esperado], um jogador não escolhe, como anteriormente, sua estratégia [ou seja, um vértice do triângulo, uma estratégia pura], mas joga todas as estratégias possíveis e escolhe apenas as probabilidades com as quais irá jogá-las, respectivamente. [...] vimos que a característica desse caso é que constitui uma desvantagem definitiva para cada jogador ter suas intenções descobertas pelo oponente. Assim, uma consideração importante para um jogador é proteger-se contra a descoberta de suas intenções pelo oponente. **Jogar várias estratégias diferentes aleatoriamente, de modo que apenas suas probabilidades sejam determinadas, é uma maneira muito eficaz de obter certo grau de proteção: com esse dispositivo, o oponente não pode descobrir qual será a estratégia do jogador, já que ele mesmo não sabe [o que jogará]. A ignorância é obviamente uma salvaguarda muito boa contra a divulgação de informações direta ou indiretamente** (grifo e tradução do autor).²¹

Nessa situação, em que o segredo se torna fundamental, o jogador utiliza uma estratégia mista para determinar como escolherá uma linha ou coluna. Por exemplo, se sua estratégia mista é $r = (\frac{2}{5}, \frac{3}{5})$, digamos, $\frac{2}{5}$ para a linha um e $\frac{3}{5}$ para a linha dois, então o jogador pode quebrar dois palitos e juntar no bolso com mais três palitos inteiros, pôr a mão no bolso e pegar o primeiro palito que tocar com os dedos. Se um palito quebrado for selecionado, escolhe a linha um; se selecionar um palito inteiro, joga a linha dois. Esse é o sentido que deve ser atribuído às estratégias mistas.

3.2 Determinando selas

Vejamos como procedemos em geral. Seja uma matriz

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \neq 0,$$

em que $a_{ij} \geq 0$. Essa matriz representa os *payoffs* do jogador das linhas. Este jogador escolhe $r = (r_1, \dots, r_m) \in \Delta_m = \{(r_1, \dots, r_m) \in \mathbb{R}^m \mid r_1 + \dots + r_m = 1, r_i \geq 0\}$. O jogador das colunas, por sua vez, escolhe $c = (c_1, \dots, c_n) \in \Delta_n = \{(c_1, \dots, c_n) \in \mathbb{R}^n \mid c_1 + \dots + c_n = 1, c_j \geq 0\}$. A função $E(r, c)$ é definida como:

$$E(r, c) = r' \cdot A \cdot c = (r_1 \ \cdots \ r_m) \cdot \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \sum_{i,j=1}^{i=m, j=n} a_{ij} \cdot r_i \cdot c_j.$$

Vamos determinar a solução do jogador das colunas. Como $E(e_i, c^*) \leq E(r^*, c^*) = v^*, i = 1, \dots, m$, então:

$$\begin{aligned} a_{11} \cdot c_1^* + \cdots + a_{1n} \cdot c_n^* &\leq v^* \\ &\vdots \\ a_{m1} \cdot c_1^* + \cdots + a_{mn} \cdot c_n^* &\leq v^*. \end{aligned}$$

Vamos supor que $v^* > 0$ e isso, como comentaremos mais à frente, não é uma limitação para o procedimento. Então:

$$\begin{aligned} a_{11} \cdot \left(\frac{c_1^*}{v^*}\right) + \cdots + a_{1n} \cdot \left(\frac{c_n^*}{v^*}\right) &\leq 1 \\ &\vdots \\ a_{m1} \cdot \left(\frac{c_1^*}{v^*}\right) + \cdots + a_{mn} \cdot \left(\frac{c_n^*}{v^*}\right) &\leq 1. \end{aligned}$$

Substituindo $\frac{c_j^*}{v^*} = x_j^*, j = 1, \dots, n$, resulta

$$\begin{aligned} a_{11} \cdot x_1^* + \cdots + a_{1n} \cdot x_n^* &\leq 1 \\ &\vdots \\ a_{m1} \cdot x_1^* + \cdots + a_{mn} \cdot x_n^* &\leq 1, \end{aligned}$$

em que $\frac{c_1^*}{v^*} + \cdots + \frac{c_n^*}{v^*} = \frac{c_1^* + \cdots + c_n^*}{v^*} = \frac{1}{v^*} = x_1^* + \cdots + x_n^*$. Como o objetivo do jogador das colunas é minimizar $E(r^*, c)$, então v^* deve maximizar $\frac{1}{v^*}$. Portanto, o problema do jogador das colunas se expressa no problema primal

$$\begin{aligned} \max \quad & x_1 + \cdots + x_n \\ \text{sa} \quad & a_{11} \cdot x_1^* + \cdots + a_{1n} \cdot x_n^* \leq 1 \\ & \vdots \qquad \qquad \qquad \vdots \\ & a_{m1} \cdot x_1^* + \cdots + a_{mn} \cdot x_n^* \leq 1 \\ & x_1 \geq 0, \dots, x_n \geq 0. \end{aligned}$$

Vamos determinar a solução do jogador das linhas. Como $E(r^*, e_j) \geq E(r^*, c^*) \leq v^*, j=1, \dots, n$, então:

$$\begin{aligned} a_{11} \cdot r_1^* + \dots + a_{m1} \cdot r_m^* &\geq v^* \\ &\vdots \\ a_{1n} \cdot r_1^* + \dots + a_{mn} \cdot r_m^* &\geq v^*. \end{aligned}$$

Como $v^* > 0$, então:

$$\begin{aligned} a_{11} \cdot \left(\frac{r_1^*}{v^*}\right) + \dots + a_{m1} \cdot \left(\frac{r_m^*}{v^*}\right) &\geq 1 \\ &\vdots \\ a_{1n} \cdot \left(\frac{r_1^*}{v^*}\right) + \dots + a_{mn} \cdot \left(\frac{r_m^*}{v^*}\right) &\geq 1. \end{aligned}$$

Substituindo $\frac{r_i^*}{v^*} = y_i^*, i=1, \dots, m$, resulta:

$$\begin{aligned} a_{11} \cdot y_1^* + \dots + a_{m1} \cdot y_m^* &\geq 1 \\ &\vdots \\ a_{1n} \cdot y_1^* + \dots + a_{mn} \cdot y_m^* &\geq 1, \end{aligned}$$

em que $\frac{r_1^*}{v^*} + \dots + \frac{r_m^*}{v^*} = \frac{r_1^* + \dots + r_m^*}{v^*} = \frac{1}{v^*} = y_1^* + \dots + y_m^*$. Como o objetivo do jogador das linhas é maximizar $E(r, c^*)$, então deve minimizar $\frac{1}{v^*}$. Portanto, o problema do jogador das linhas se expressa no problema dual

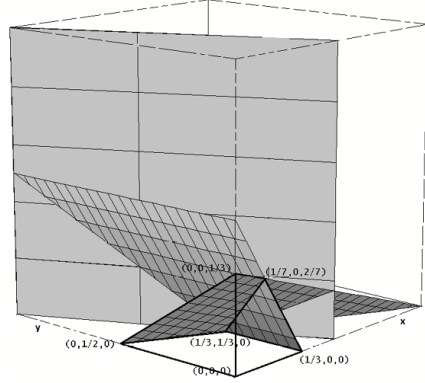
$$\begin{aligned} \min \quad & y_1 + \dots + y_m \\ \text{sa} \quad & a_{11} \cdot y_1^* + \dots + a_{m1} \cdot y_m^* \geq 1 \\ & \vdots \quad \quad \quad \vdots \\ & a_{1n} \cdot y_1^* + \dots + a_{mn} \cdot y_m^* \geq 1 \\ & y_1 \geq 0, \dots, y_m \geq 0. \end{aligned}$$

Exemplo 12: Consideremos o jogo $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 3 \\ 3 & 0 & 2 \end{pmatrix}$. O problema do jogador das colunas é:

$$\begin{aligned} \max \quad & x_1 + x_2 + x_3 \\ \text{sa} \quad & 2 \cdot x_1 + x_2 + 0 \cdot x_3 \leq 1 \\ & x_1 + 2 \cdot x_2 + 3 \cdot x_3 \leq 1 \\ & 3 \cdot x_1 + 0 \cdot x_2 + 2 \cdot x_3 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{aligned}$$

Veja o gráfico abaixo. Os vértices da restrição são: $(0,0,0)$, $(\frac{1}{3}, \frac{1}{3}, 0)$, $(\frac{1}{3}, 0, 0)$, $(0, \frac{1}{2}, 0)$, $(0, 0, \frac{1}{3})$ e $(\frac{1}{7}, 0, \frac{2}{7})$. O máximo do problema acima está no vértice

$(\frac{1}{3}, \frac{1}{3}, 0)$, em que a soma $x_1 + x_2 + x_3 = \frac{2}{3}$. O valor do jogo é $v = \frac{3}{2}$ e a estratégia ótima para o jogador das colunas é $c = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$.



O problema do jogador das linhas é:

$$\begin{array}{ll} \min & y_1 + y_2 + y_3 \\ \text{sa} & 2 \cdot y_1 + y_2 + 3 \cdot y_3 \geq 1 \\ & y_1 + 2 \cdot y_2 + 0 \cdot y_3 \geq 1 \\ & 0 \cdot y_1 + 3 \cdot y_2 + 2 \cdot y_3 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

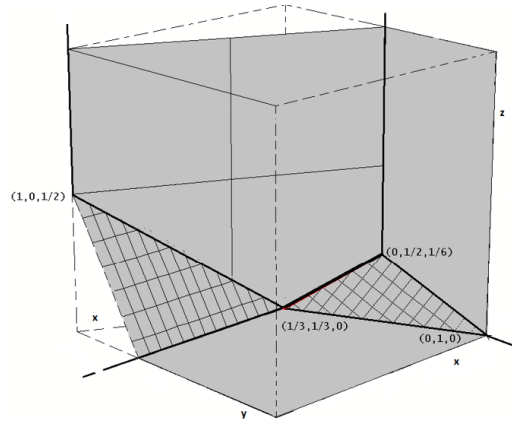
Veja a figura abaixo. Os vértices da região determinada pela restrição são: $(1, 0, \frac{1}{2})$, $(\frac{1}{3}, \frac{1}{3}, 0)$, $(0, \frac{1}{2}, \frac{1}{6})$ e $(0, 1, 0)$. O mínimo está em dois vértices: $(\frac{1}{3}, \frac{1}{3}, 0)$ e $(0, \frac{1}{2}, \frac{1}{6})$. Portanto, existem pelo menos duas estratégias

ótimas que são: $r = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$ e $r = \begin{pmatrix} 0 \\ 3/4 \\ 1/4 \end{pmatrix}$. Mas, são as únicas estratégias ótimas

para o jogador das linhas, pois todo o segmento de reta entre os pontos $(\frac{1}{3}, \frac{1}{3}, 0)$ e $(0, \frac{1}{2}, \frac{1}{6})$, ou seja, $(\frac{1}{3}, \frac{1}{3}, 0) + t((0, \frac{1}{2}, \frac{1}{6}) - (\frac{1}{3}, \frac{1}{3}, 0))$, para $0 \leq t \leq 1$, isto é,

$y_1 = \frac{1}{3} - \frac{t}{3}$, $y_2 = \frac{1}{3} + \frac{t}{6}$, $y_3 = 0 + \frac{t}{6}$, para $0 \leq t \leq 1$, são tais que $y_1 + y_2 + y_3 = \frac{2}{3}$, o que implica que as

estratégias ótimas para o jogador das linhas são $r = \begin{pmatrix} 1-t/2 \\ 2+t/4 \\ t/4 \end{pmatrix}$, $0 \leq t \leq 1$.



Façamos a contas para verificar que as estratégias $r^* = \begin{pmatrix} \frac{1-t}{2} \\ \frac{2+t}{4} \\ \frac{t}{4} \end{pmatrix}$, $0 \leq t \leq 1$ e $c^* = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$ são selas do *payoff* esperado:

$$r'' \cdot A \cdot c^* = \begin{pmatrix} \frac{1-t}{2} & \frac{2+t}{4} & \frac{t}{4} \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 3 \\ 3 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} = \frac{3}{2};$$

$$r' \cdot A \cdot c^* = \begin{pmatrix} r_1 & r_2 & r_3 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 3 \\ 3 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} = \frac{3}{2} \cdot (r_1 + r_2 + r_3) = \frac{3}{2};$$

$$r' \cdot A \cdot c = \begin{pmatrix} \frac{1-t}{2} & \frac{2+t}{4} & \frac{t}{4} \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 3 \\ 3 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \frac{3}{2} \cdot (c_1 + c_2 + c_3) + \frac{5 \cdot t \cdot c_3}{4} \geq \frac{3}{2},$$

portanto, $r' \cdot A \cdot c^* = r'' \cdot A \cdot c^* \leq r' \cdot A \cdot c$, mostrando que (r^*, c^*) é uma sela.

Exemplo 13: Consideremos o jogo $A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & -1 & 3 \\ -1 & 2 & -1 \end{pmatrix}$, que foi retirado das páginas 43-44 do livro de McKinsey²². Como veremos, o valor desse jogo não é positivo e, portanto, não podemos ratá-lo exatamente como fizemos anteriormente. Se ocorrer de o valor v^* ser nulo, não podemos dividir por v^* como fizemos; se ocorrer de o valor ser negativo, as desigualdades se invertem. Entretanto, se somarmos aos *payoffs* um valor positivo qualquer, estaremos alterando o valor do jogo, mas não as estratégias adotadas pelos jogadores. No caso

acima, podemos somar 1 e obter um novo jogo $B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 3 & 0 \end{pmatrix}$. O problema do jogador das colunas associado a este jogo é:

$$\begin{array}{ll} \max & x_1 + x_2 + x_3 \\ \text{sa} & 2 \cdot x_1 \leq 1 \\ & 4 \cdot x_3 \leq 1 \\ & 3 \cdot x_2 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{array}$$

O problema do jogador das linhas é:

$$\begin{array}{ll} \min & y_1 + y_2 + y_3 \\ \text{sa} & 2 \cdot y_1 \geq 1 \\ & 3 \cdot y_3 \geq 1 \\ & 4 \cdot y_2 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

Ambos os problemas são fáceis de resolver, obviamente $x^* = \begin{pmatrix} 1/2 \\ 1/3 \\ 1/4 \end{pmatrix}$ e $y^* = \begin{pmatrix} 1/2 \\ 1/4 \\ 1/3 \end{pmatrix}$. O valor do jogo B é tal que $\frac{1}{v^*} = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}$, ou seja, $v^* = \frac{12}{13}$. As estratégias ótimas são $r^* = \begin{pmatrix} 6/13 \\ 3/13 \\ 4/13 \end{pmatrix}$ e $c^* = \begin{pmatrix} 6/13 \\ 4/13 \\ 3/13 \end{pmatrix}$. Essas também são as estratégias ótimas para o jogo A , cujo o valor é $v^* = \frac{12}{13} - 1 = -\frac{1}{13}$.

Exercício 5

Considere o jogo:

$$\begin{pmatrix} 3 & 2 & 4 \\ 1 & 4 & 0 \end{pmatrix}.$$

Formule os problemas lineares (sem resolvê-los) para o jogador das linhas e para o jogador das colunas.

Exercício 6

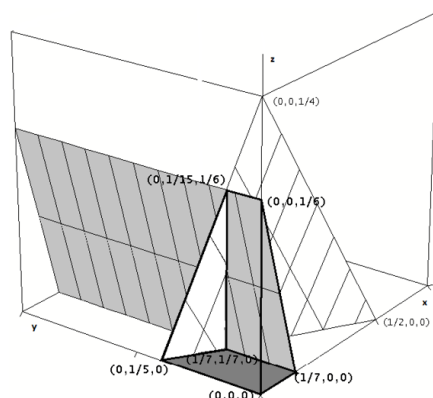
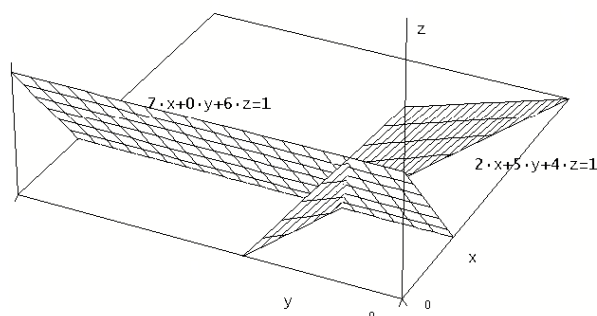
Suponhamos que o jogo seja dado pela matriz $A = \begin{pmatrix} -2 & 1 & 0 \\ 3 & -4 & 2 \end{pmatrix}$.

Não é difícil perceber que se somarmos um número fixo ε aos *payoffs* a_{ij} então estamos somando ε ao *payoff* esperado $E(r, c)$, pois

$(r_1 \ r_2) \cdot \begin{pmatrix} \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \varepsilon$, para $r \in \Delta_2$ e $c \in \Delta_3$; portanto, $r' \cdot (A + \varepsilon) \cdot c = r' \cdot A \cdot c + \varepsilon$, logo o valor do

jogo v^* passa a ser $v^* + \varepsilon$, mas as estratégias ótimas r^* e c^* não se alteram. Sendo assim, para obter uma matriz $B \geq 0, B \neq 0$ somamos algum número $\varepsilon > 0$ a cada elemento da matriz A , de maneira que $A + \varepsilon \geq 0$. No exemplo, somamos $\varepsilon = 4$ e obtemos $B = \begin{pmatrix} 2 & 5 & 4 \\ 7 & 0 & 6 \end{pmatrix}$. O valor do jogo original está somado de +4 e as estratégias ótimas permanecem as mesmas, tanto no jogo definido por A quanto no jogo definido por B . Escreva os problemas primal e dual que determinam as estratégias ótimas para os dois jogadores em relação à matriz A usando a matriz B . Resolva os problemas dos jogadores das colunas e das linhas e respectivas estratégias ótimas. Use os gráficos abaixo.

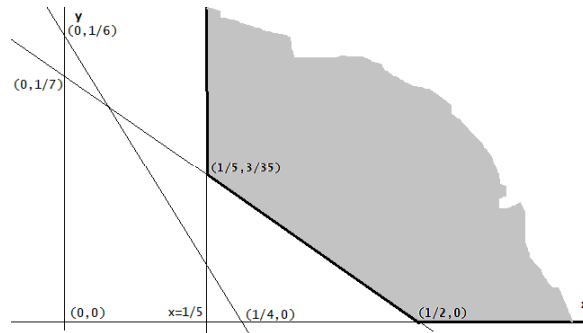
Visualize abaixo os planos que determinam a restrição do problema do jogador da coluna. A seguir está o mesmo gráfico com os vértices da região determinada pelas restrições:



Verifique que o máximo da soma $x_1 + x_2 + x_3 = \frac{2}{7}$ está no vértice $(\frac{1}{7}, \frac{1}{7}, 0)$. As estratégias ótimas para o jogador das colunas é $c^* = (\frac{1}{7}, \frac{1}{7}, 0)$; o valor do jogo

definido pela matriz B é e o valor do jogo definido pela matriz A é .

Abaixo está a restrição do jogador das linhas. Certamente, o mínimo de (por que?) no vértice ; a estratégia ótima para o jogador das linhas é .



Verifique que , $\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \cdot \begin{pmatrix} 2 & 5 & 4 \\ 7 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} = \frac{7}{2} \cdot (r_1 + r_2)$;

$\begin{pmatrix} 7/10 \\ 3/10 \end{pmatrix} \cdot \begin{pmatrix} 2 & 5 & 4 \\ 7 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \frac{7 \cdot c_1}{2} + \frac{7 \cdot c_2}{2} + \frac{23 \cdot c_3}{5}$; conclua que $r^* = \begin{pmatrix} 7/10 \\ 3/10 \end{pmatrix}$ e $c^* = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$ é uma sela para o *payoff* esperado $\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \cdot \begin{pmatrix} 2 & 5 & 4 \\ 7 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$.

Verifique que $\begin{pmatrix} 7/10 \\ 3/10 \end{pmatrix} \cdot \begin{pmatrix} -2 & 1 & 0 \\ 3 & -4 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} = -\frac{1}{2}$, $\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \cdot \begin{pmatrix} 2 & 5 & 4 \\ 7 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} = -\frac{(r_1 + r_2)}{2}$;

$\begin{pmatrix} 7/10 \\ 3/10 \end{pmatrix} \cdot \begin{pmatrix} -2 & 1 & 0 \\ 3 & -4 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = -\frac{c_1}{2} - \frac{c_2}{2} + \frac{23 \cdot c_3}{5}$; conclua que e $c^* = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$ é uma sela para o *payoff* esperado $\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \cdot \begin{pmatrix} -2 & 1 & 0 \\ 3 & -4 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$.

3.4 A Teoria dos Jogos vai à guerra (Exemplo 14)

Oliver Garfield Haywood Jr. (1911 –2002) foi um militar norte-americano que alcançou a patente de brigadeiro. Haywood formou-se na academia militar de West Point como o primeiro da turma e doutorou-se em Engenharia Civil pelo MIT em 1940. Depois da

guerra, Haywood integrou o projeto Manhattan, tendo trabalhado com Edward Teller (1908-2003), húngaro naturalizado norte-americano, idealizador da bomba de hidrogênio, que dizia que “o coronel Haywood é o único militar para quem eu trabalharia”, segundo o *The Manhattan Project Heritage Preservation Association, Inc.* Lá também está registrado que Haywood “comentou que pensava ser uma das poucas pessoas no mundo que conseguia projetar uma arma termonuclear inteira”²³.

Neste exemplo, analisaremos o artigo “*Military Decision and Game Theory*” (1954), de Haywood Jr. Para ele,

um comandante militar tem duas filosofias para auxiliá-lo na decisão. Ele pode selecionar seu curso de ação com base em sua avaliação sobre o que seu inimigo tem condições de fazer para se opor ou ele pode fazer sua seleção com base em sua suposição sobre o que inimigo fará. A primeira filosofia é uma doutrina de decisão baseada nas capacidades inimigas; a última, baseada nas intenções inimigas. A doutrina da decisão das forças armadas dos Estados Unidos é uma doutrina baseada nas capacidades inimigas.²⁴

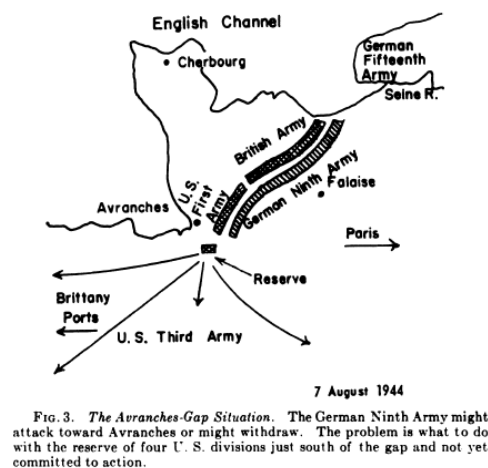
Como exemplo, Haywood analisa uma batalha ocorrida em 1944, durante o desembarque das forças aliadas na Europa. No comando dos norte-americanos estava o general Omar Bradley, que tinha que tomar uma decisão em relação às divisões de reserva de que dispunha. Ao Oeste, na proximidade da cidade francesa de Avranches, existia a possibilidade de acesso marítimo (brecha) para o exército alemão comandado pelo general von Kluge. Bradley considerou três estratégias:

- 1) Conduzir as forças de reserva para defender a brecha dificultando o acesso ao mar por parte do nono exército alemão;
- 2) Conduzir as forças de reserva para o Leste, para fustigar a retaguarda do exército alemão caso este, buscando ocupar a brecha, optasse por marchar para o Oeste, atacando as forças

aliadas constituídas pelo primeiro exército norte-americano e pelo exército britânico postadas à frente; caso o exército alemão recuasse, as forças de reserva poderiam ser utilizadas para perseguir o exército alemão em direção ao Leste;

- 3) Aguardar um dia para observar as ações alemães. Caso o exército alemão optasse por atacar e as forças aliadas resistissem ao ataque mantendo a brecha, então Bradley conduziria as forças de reserva para o Oeste cercando os alemães; se o nono exército alemão optasse pelo recuo, as forças de reserva se deslocariam para o Leste, atacando a retaguarda das forças alemães em fuga.

Veja o croqui abaixo retirado do artigo do Haywood²⁵



Para o nono exército do general von Kluge, existiam duas estratégias: atacar em direção ao Oeste ou retirar-se em direção ao Leste. As três estratégias disponíveis para Bradley e as duas disponíveis para von Kluge admitem seis resultados ou “outcomes”. Hayhood elaborou um quadro com os resultados:

	von Kluge ataca	von Kluge recua
Bradley protege a	Bradley desloca as reservas para Oeste para reforçar a brecha; von Kluge ataca.	Bradley desloca as reservas para Oeste para reforçar a brecha; von Kluge recua.

brecha no Oeste	Resultado estimado: as forças aliadas repelem o ataque e a brecha permanece inacessível para os alemães.	Resultado estimado: Bradley exerce uma pressão fraca sobre a retaguarda alemã.
Bradley desloca-se para o Leste	Bradley desloca as reservas para o Leste; Von Kluge ataca. Resultado estimado: os alemães conquistam a brecha	Bradley desloca as reservas para o Leste; von Kluge recua. Resultado estimado: Bradley exerce forte pressão sobre a retirada alemã.
Bradley aguarda	Bradley aguarda um dia para observar as ações alemães; se von Kluge optar pelo ataque, Bradley dirige as forças de reserva para o Oeste. Resultado estimado: a brecha se mantém e o nono exército alemão é cercado.	Bradley aguarda um dia para observar as ações alemães; se von Kluge optar pelo recuo, Bradley dirige as forças de reserva para o Leste. Resultado estimado: pressão moderada sobre a retaguarda alemã.

A seguir, Bradley tem que estabelecer uma forma de comparar esses resultados, Hayhood observa: “Antes de poder julgar o que era melhor e o que era pior, ele [Bradley] teve que estabelecer sua ordem de preferência para estes resultados”²⁶. A ordem de preferência, do melhor resultado para o pior foi:

- A. A brecha é mantida, possivelmente alemães cercados;
- B. Forte pressão sobre a retirada alemã;
- C. Pressão moderada sobre a retirada alemã;
- D. Pressão fraca na retirada alemã;
- E. A brecha é mantida;
- F. Os alemães conquistam a brecha.

Portanto, temos , em que significa estritamente preferível.

Acima, o quadro de Hayhood²⁷ para explicar os resultados estimados por Bradley. Temos que supor, também, que o general alemão avalie os resultados de forma semelhante, obviamente invertendo a ordem de preferência. Para von Kluge, .

Se representarmos o jogo acima numa tabela, temos:

.

Não temos números que correspondam aos *payoffs* dos jogadores, pois apenas estabelecemos uma ordem de preferências para os possíveis resultados, mas, mesmo assim, vamos conduzir a nossa análise utilizando a técnica minimax de von Neumann. Em cada linha que corresponde às escolhas do general Bradley, os resultados menos preferíveis são $\{E, F, C\}$, ou seja, os níveis de segurança de Bradley; entre esses resultados, o preferível é c . O resultado C é o *payoff* de Bradley, por adotar a estratégia maxmin. Para o general von Kluge, os piores resultados são $\{A, B\}$, pois von Kluge escolhe nas colunas; o melhor resultado para Kluge entre A e B é B . Esse é resultado minmax para o general von Kluge. Como $B \succ C$, então não existe um *payoff* comum para ambos e, portanto, não existe sela em estratégias puras para o *payoff* esperado do general Bradley. Observe que se o jogo tem uma sela em estratégias puras, então $B \approx C$, isto é, os generais seriam indiferentes entre os resultados B e C .

Se escolhêssemos números em lugar de letras não teríamos resultado diferente. Suponhamos que a ordem de preferência possa ser representada por uma utilidade, ou seja, simplesmente, atribuímos números compatíveis com o ordenamento das preferências aos diferentes resultados. Por exemplo, seja o jogo $M = \begin{pmatrix} 2 & 3 \\ 1 & 5 \\ 6 & 4 \end{pmatrix}$. A estratégia minmax para o jogador das linhas é o ; a estratégia maxmin para o jogador das colunas é o . Esse jogo não tem solução

em estratégias puras, pois o valor . Logo, teremos de procurar selas em estratégias mistas.

Em geral, como já comentamos:

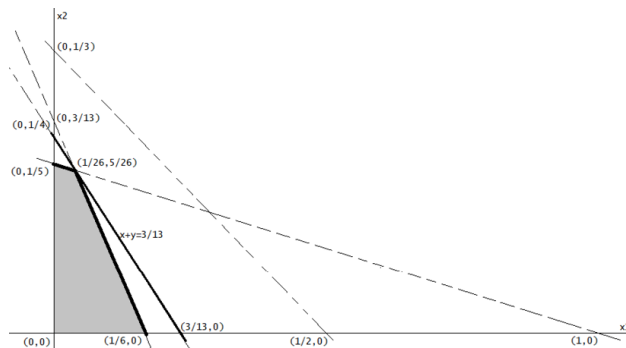
.

Von Neumann observou esse resultado no artigo “*On The Theory of Games of Strategy*”²⁸. De fato, mostrar que existe sela para os *payoffs* esperados é equivalente a mostrar que . Por quê?

Voltemos ao jogo em que atribuímos números aos resultados. A estratégia ótima para von Kluge é dada pelo problema:

$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{sa} & 2 \cdot x_1 + 3 \cdot x_2 \leq 1 \\ & x_1 + 5 \cdot x_2 \leq 1 \\ & 6 \cdot x_1 + 4 \cdot x_2 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Veja o gráfico abaixo. O valor deste jogo é $v = \frac{13}{3}$ e as estratégias ótimas para von Kluge são $c = \left(\frac{1}{6}, \frac{5}{6}\right)$. Observe que se este fosse o jogo em que os dois generais se enfrentam, então as chances que von Kluge recue são de cinco para uma.



Digamos que trabalhamos com nossa matriz original e que traduzimos a tabela de preferências de Bradley em números $A > B > C > D > E > F$. O resultado não será muito diferente. A estratégia ótima para von Kluge é dada pelo problema:

$$\begin{array}{ll}
\max & x_1 + x_2 \\
\text{sa} & E \cdot x_1 + D \cdot x_2 \leq 1 \\
& F \cdot x_1 + B \cdot x_2 \leq 1 \\
& A \cdot x_1 + C \cdot x_2 \leq 1 \\
& x_1 \geq 0, x_2 \geq 0.
\end{array}$$

Observando que $\frac{1}{A} < \frac{1}{B} < \frac{1}{C} < \frac{1}{D} < \frac{1}{E} < \frac{1}{F}$, os vértices da região viável são $v = \left(\frac{1}{A}, 0\right)$, $\left(\frac{B-C}{A \cdot B - C \cdot F}, \frac{A-F}{A \cdot B - C \cdot F}\right)$ e $\left(0, \frac{1}{B}\right)$; a solução do problema está no vértice $\left(\frac{B-C}{A \cdot B - C \cdot F}, \frac{A-F}{A \cdot B - C \cdot F}\right)$, porque $\frac{A}{C} > 1 > \frac{F}{B}$; o valor do jogo é $v = \frac{A \cdot B - C \cdot F}{(A+B) - (C+F)}$. A estratégia ótima para o general von Kluge é:

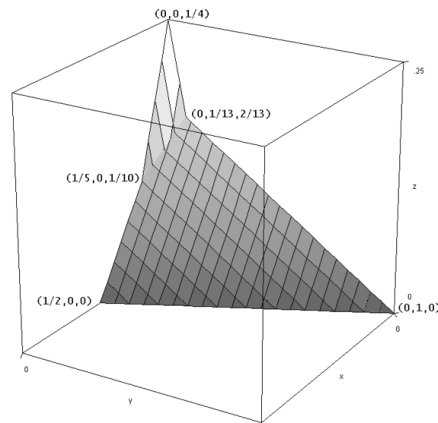
$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \frac{B-C}{(A+B)-(C+F)} \\ \frac{A-F}{(A+B)-(C+F)} \end{pmatrix}.$$

Observe que $c_2 = \frac{A-F}{(A+B)-(C+F)} > \frac{B-C}{(A+B)-(C+F)} = c_1$. Portanto, von Kluge sempre estará mais disposto a optar pela retirada, independentemente dos valores numéricos atribuídos ao ordenamento das preferências. De fato, von Kluge quis recuar, mas Hitler o proibiu!

A estratégia ótima para Bradley é dada pelo problema:

$$\begin{array}{ll}
\min & y_1 + y_2 + y_3 \\
\text{sa} & 2 \cdot y_1 + y_2 + 6 \cdot y_3 \geq 1 \\
& 3 \cdot y_1 + 5 \cdot y_2 + 4 \cdot y_3 \geq 1 \\
& y_1 \geq 0, y_2 \geq 0, y_3 \geq 0.
\end{array}$$

Os vértices da região determinada pelas restrições estão abaixo:



O mínimo está no vértice $(0, \frac{1}{13}, \frac{2}{13})$, obviamente. A estratégia ótima para Bradley é $r = \begin{pmatrix} 0 \\ 1/3 \\ 2/3 \end{pmatrix}$. Se o jogador das linhas fosse o general Bradley, ele escolheria esperar um dia antes de tomar uma decisão em relação a deslocar as reservas para Leste numa proporção de dois para um. De fato, Bradley optou por aguardar.

O fato histórico foi que Bradley aguardou um dia para ver como os alemães agiriam. Coagido por Hitler, von Kluge avançou. As forças americanas, inglesas e francesas resistiram ao ataque e Bradley, no segundo dia, atacou o exército alemão em retirada. O general alemão se suicidou! Como dizem Luce e Rice:

imagine uma investigação do congresso sobre um comandante militar ou um chefe de agência que adotou uma estratégia pura e específica ruinosa. Qual seria a reação se sua defesa dependesse do fato de que ele adotou essa estratégia pura através do lançamento de um dado? Ou igualmente, imagine um comandante tão mal aconselhado que, ao ser elogiado por uma brilhante e bem-sucedida jogada estratégica, revela que sua escolha resultou do lançamento de uma moeda. Infelizmente, o estrategista é frequentemente avaliado em termos do resultado da escolha adotada, e não em termos de sua conveniência estratégica em toda a situação de risco²⁹.

Exercício 7

Novamente, suponhamos que não tenhamos valores para os *payoffs* e façamos a análise do jogo com . Verifique que o problema do Bradley é:

$$\begin{array}{ll} \min & y_1 + y_2 + y_3 \\ \text{sa} & E \cdot y_1 + F \cdot y_2 + A \cdot y_3 \geq 1 \\ & D \cdot y_1 + B \cdot y_2 + C \cdot y_3 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

i) Encontre os vértices da região determinada pelas restrições que são $(\frac{1}{E}, 0, 0)$, $(0, \frac{1}{F}, 0)$, $(0, 0, \frac{1}{C})$, $(0, \frac{A-C}{A \cdot B - C \cdot F}, \frac{B-F}{A \cdot B - C \cdot F})$ e $(\frac{A-C}{A \cdot D - C \cdot E}, 0, \frac{D-E}{A \cdot D - C \cdot E})$.

ii) Mostre que a solução do problema acima está no vértice $(0, \frac{A-C}{A \cdot B - C \cdot F}, \frac{B-F}{A \cdot B - C \cdot F})$. Não é necessário calcular a soma $y_1 + y_2 + y_3$ nos cinco

vértices, pois já sabemos que o valor do jogo é $v = \frac{A \cdot B - C \cdot F}{(A+B)-(C+F)}$.

iii) Obtenha a estratégia ótima para o general Bradley

$r = \begin{pmatrix} 0 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{A-C}{(A+B)-(C+F)} \\ \frac{B-F}{(A+B)-(C+F)} \end{pmatrix}$. Entretanto, nesse caso, se $r_2 < r_3$ ou $r_2 = r_3$ ou $r_2 > r_3$, depende

dos valores escolhidos para A, B, C, D, E e F . Portanto, Bradley teve que decidir entre duas opções: dirigir as forças para o Leste ou aguardar. Optou por aguardar.

Sempre que mantivermos a distância entre esses números constante, isto é, $A-B=B-C=C-D=D-E=E-F=\Delta$, então $r_2 < r_3$. Esse resultado é compatível com a decisão de Bradley. Mas, se Bradley é um general avesso ao risco, é razoável supormos que $A-C < B-F$ o que não implica que a escolha de Bradley sempre recairá sobre a estratégia três.

Exercício 8

Digamos que o General Bradley seja indiferente entre exercer uma pressão forte sobre a retirada alemã e cercar os alemães mantendo a brecha. Poderíamos representar o jogo por $M = \begin{pmatrix} 2 & 3 \\ 1 & 5 \\ 5 & 4 \end{pmatrix}$. Formule o problema para o jogador das colunas e encontre as estratégias ótimas para este jogador. Resposta: $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \\ \frac{4}{5} \end{pmatrix}$.

i) Formule o problema para o jogador das linhas e encontre as estratégias ótimas para este jogador. Resposta: $r = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/5 \\ 4/5 \end{pmatrix}$.

ii) Observe que o valor do jogo é $v = \frac{21}{5} = 4,2$.

Se em segredo, obviamente, Bradley decide que optou pela estratégia três e supondo que von Kluge não tem como saber dessa decisão (ou seja, seus informantes não souberam da decisão de Bradley), então o *payoff* esperado por Bradley é $E = 5 \cdot \frac{1}{5} + 4 \cdot \frac{4}{5} = \frac{21}{5}$, pois Kluge

manteria sua estratégia $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \\ \frac{4}{5} \end{pmatrix}$. Entretanto, se Kluge souber dessa escolha, certamente optará pela estratégia $c = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e o *payoff* de Bradley será quatro.

Haywood comenta que o exército norte-americano segue a doutrina da “Estimativa de Situação” que é

Inquestionavelmente, conservadora; von Neumann prova com sua matemática difícil que essa doutrina constitui a filosofia de decisão mais conservadora. De fato, um comandante é obrigado a assumir que o inimigo pode descobrir sua decisão e adotar a estratégia mais eficaz em oposição. Assim, a doutrina dos EUA minimiza as consequências indesejáveis que resultam do inimigo descobrir ou adivinhar a estratégia dos EUA. O risco é reduzido, no entanto, à custa de uma possível exploração de um palpite correto sobre as intenções do inimigo. Uma doutrina que baseie a decisão em uma estimativa do que o inimigo pretende fazer, e não do que ele é capaz de fazer, é inquestionavelmente uma doutrina mais ousada, mas mais arriscada.³⁰

Haywood observa que, eventualmente, os comandantes americanos se desviaram da doutrina padrão, seja porque possuíam relatórios da inteligência sobre quais seriam as ações inimigas, seja porque conheciam o comportamento dos oponentes ou tinham uma compreensão (ou pensavam ter) das intenções do inimigo:

A liderança e a impetuosidade do general Patton no campo de batalha eram em parte devidos a um sentimento de compreensão das intenções do inimigo. Seus avanços aparentemente imprudentes foram possíveis graças à sua confiança que ele poderia se antecipar ao ataque inimigo e preparar suas forças para recebê-lo. A fama alcançada pelo general Michaelis no primeiro ano da luta coreana repousou em grande parte em sua capacidade de estimar o padrão de pensamento de combate dos generais da Coreia do Norte.

Mas,

Por outro lado, Pearl Harbor é um exemplo clássico de desastre através de uma decisão baseada em uma estimativa errada das intenções do inimigo. Antes do ataque, não havia dúvida de que os japoneses eram capazes de lançar aeronaves de porta-aviões contra Pearl Harbor. Essa capacidade foi ignorada na crença firme de que os japoneses atacariam primeiro com sabotadores e submarinos. O resultado foi uma movimentação

incorreta das forças dos EUA - uma movimentação contra uma estimativa do que os japoneses pretendiam fazer, não contra o que eles eram capazes de fazer.³¹

Defendendo, a doutrina da avaliação da situação, Haywood afirma em relação a um comandante militar:

se sua avaliação estiver correta, ele obtém certa segurança baseando sua decisão nas capacidades de seu inimigo. Se o inimigo errar em sua avaliação ou tomar uma decisão estúpida, não poderá colocar o comandante em uma posição menos favorável do que aquela que ele havia previsto. Na medida em que uma doutrina de decisão baseada nas capacidades inimigas é conservadora, um comandante que estimou a situação corretamente protege suas forças da surpresa ou dos danos causados pelos erros do inimigo.

E, alerta:

Esse não é o caso de uma doutrina baseada em uma estimativa das intenções do inimigo. Um comandante pode enfrentar um desastre não por causa de seu erro, mas por causa do erro de seu inimigo. Ele pode se encontrar incorretamente mobilizado para a batalha porque o inimigo analisou a situação de forma inadequada ou tomou uma decisão errada. Obviamente, ele também pode achar que o inimigo seja mais sábio do que ele e adivinhar suas intenções. De qualquer forma, um comandante corre o risco de se melindrar se ignorar a capacidade inimiga de adivinhar as intenções do oponente. Mas se ele adivinhar corretamente, ele sempre pode obter um desdobramento da batalha pelo menos tão favorável quanto o melhor resultado que se poderia garantir com a doutrina baseada nas capacidades inimigas (Essa afirmação é comprovada matematicamente na teoria dos jogos). Existe uma promessa de ganho sob risco, mas o comandante deve estar certo em sua decisão.³²

Em relação a esse alerta, podemos comentar que se Bradley tivesse baseado sua decisão nas intenções do general alemão (por exemplo, se os informantes alertassem Bradley sobre a intenção de von Kluge, mas não soubessem que o recuo se tornara impraticável por causa das ordens de Hitler), ele concluiria que poderia fustigar intensamente a retirada alemã. Todavia, Bradley se exporia à possibilidade que os alemães conquistassem o acesso ao mar, o que seria o pior resultado para os Aliados (resultado F). Como de fato os alemães atacaram, o resultado F tornou-se uma possibilidade real.

Os comandantes aliados lamentariam o resultado, atribuindo a ousadia de von Kluge ao possível conhecimento que teve das intenções de Bradley e à sua inabilidade! Na verdade, von Kluge estava apenas cedendo à avaliação delirante de Hitler!

Exercício 9 – Batalha do Mar de Bismark

Outro exemplo que Haywood analisou foi uma batalha ocorrida em 1943 na Nova Guiné. Um brigadeiro norte-americano chamado Kenney recebeu ordens de interceptar e causar o máximo de dano possível ao comboio japonês que levava tropas e suprimentos para a proximidade da cidade de Lae a partir da cidade Rabaul, situada numa ilha chamada Nova Bretanha. O comboio tem duas estratégias disponíveis: deslocar-se pelo Norte da ilha ou pelo Sul, consumindo, em qualquer das rotas, três dias. O sistema de previsão das condições atmosféricas alertou que o Norte da ilha estaria chuvoso e com baixa visibilidade, e que o Sul teria condições atmosféricas boas. As condições atmosféricas são importantes em relação ao desempenho aéreo.

Kenney considerou duas estratégias: enviar seu reconhecimento aéreo para o Norte ou para o Sul; uma vez localizado, o comboio poderia ser bombardeado a partir do ar.

	Comboio japonês navega pelo Norte da ilha	Comboio japonês navega pelo Sul da ilha
Kenney concentra o reconhecimento aéreo no Norte	Resultado estimado: a baixa visibilidade dificulta o reconhecimento do comboio, o que deve ocorrer no segundo dia, permitindo dois dias de bombardeio;	Resultado estimado: a concentração do reconhecimento no Norte atrasa a localização do comboio e o primeiro dia de bombardeio é desperdiçado; resulta em dois dias de bombardeio;
Kenney concentra o reconhecimento	Resultado estimado: a baixa visibilidade e a baixa	Resultado estimado: as boas condições atmosféricas e o

aéreo no Sul	concentração no Norte do	reconhecimento imediato
	reconhecimento aéreo atrasam a	permitem três dias de bombardeio.
	localização do comboio e resulta	
	em apenas um dia de	
	bombardeio.	

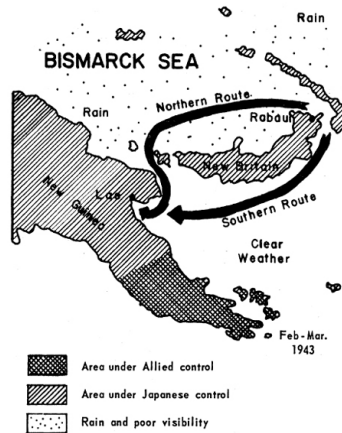
Digamos que Kenney classifica os resultados pelos dias de bombardeio, então a matriz do jogo é $A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$. Quais são as estratégias ótimas para americanos e japoneses? O problema para o jogador das colunas (comboio) é:

$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{sa} & 2 \cdot x_1 + 2 \cdot x_2 \leq 1 \\ & x_1 + 3 \cdot x_2 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0; \end{array}$$

e o problema para o jogador das linhas (força aérea) é:

$$\begin{array}{ll} \min & y_1 + y_2 \\ \text{sa} & 2 \cdot y_1 + y_2 \geq 1 \\ & y_1 + 3 \cdot y_2 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0. \end{array}$$

Resposta: $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \frac{1}{2} \leq c_1 \leq 1, c_1 + c_2 = 1$. Observe que existe uma sela em estratégias puras $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, e isso foi o que de fato ocorreu. Navegando pelo Norte da ilha, o comandante japonês se expôs a, no máximo, dois dias de bombardeio; Kenney, por sua vez, escolheu sua única estratégia maxmin e concentrou o reconhecimento no Norte. As forças japonesas foram derrotadas, mas, como observa Haywood, não se pode dizer que o comandante japonês tenha cometido um erro. Abaixo, um croqui de Haywood³³ da página 366 do artigo:



3.5 A interpretação da solução como um conjunto de probabilidades

Existe um longo trecho do livro “A Escalada do Homem”, do matemático, historiador e apresentador de televisão Jacob Bronowski (1908-1974) em que ele conta:

Trabalhei com John von Neumann na Inglaterra, durante a Segunda Guerra Mundial. Foi no interior de um táxi em Londres que ele me falou pela primeira vez sobre sua Teoria dos Jogos — esse era um dos seus locais favoritos para discussões sobre matemática. Sendo um entusiasta do xadrez logo lhe perguntei se ele se referia à teoria dos jogos afins àquela modalidade. Mas sua resposta foi negativa. “Não”, disse ele. “O xadrez não é um jogo. O xadrez é uma forma bem-definida de computação. Você pode não ser capaz de encontrar a solução correta, mas, teoricamente, sempre há uma solução, um procedimento correto para cada posição.” Agora, “jogos reais”, continuou, “em nada se assemelham a isso. A vida real não é como no xadrez. A vida real consiste em blefes, em pequenas táticas de despistamento, em perguntar a si próprio o que o parceiro está pensando sobre qual será nosso próximo movimento. É esse o tipo de jogo sobre o qual minha teoria se interessa”.³⁴ (grifo do autor)

O jogo acima, $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 2 \end{pmatrix}$, oferece soluções obviamente vantajosas para ambos os jogadores, ou seja, $r = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ e $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, em que o segredo e o blefe não impliquem que os abandonando resulte em algum ganho provável para um deles ou para ambos. Em outras palavras, nesse jogo, nenhum dos jogadores pode explorar vantajosamente um erro

do adversário e, além disso, o segredo não produz benefício para os jogadores.

Se consideramos a declaração de von Neumann, $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 2 \end{pmatrix}$ estaria excluído do conjunto dos jogos.

Exemplo 15: Dado jogo $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, utilizando o método maxmin-minmax, concluímos que $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ são estratégias ótimas para os jogadores das linhas e das colunas, respectivamente, e que $v = 1$. Observe que não existe nenhum incentivo para que o jogador das colunas escolha outra estratégia que não seja $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Não seria descabido conjecturar que o jogador das linhas, por sua vez, poderia explorar um eventual erro do jogador das colunas, adotando uma estratégia mista. Vejamos o conjunto de selas do problema.

O problema das linhas é:

$$\begin{array}{ll} \min & y_1 + y_2 \\ \text{sa} & y_1 \geq 1 \\ & y_1 + 2 \cdot y_2 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0. \end{array}$$

A única solução possível é $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. O problema do jogador das colunas é:

$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{sa} & x_1 + x_2 \leq 1 \\ & 0 \cdot x_1 + 2 \cdot x_2 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Como $v = 1$, então as estratégias disponíveis para o jogador das colunas são: $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ com $1/2 \leq c_1 \leq 1$ e $0 \leq c_2 = 1 - c_1 \leq 1/2$. Observe que $c = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ é uma estratégia mista disponível para o jogador das colunas. Na verdade, obtemos que $r^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c^* = \begin{pmatrix} c_1^* \\ c_2^* \end{pmatrix}$ com $1/2 \leq c_1^* \leq 1$ e $0 \leq c_2^* = 1 - c_1^*$ formam uma sela para E :

$$\begin{aligned} E(r, c^*) &= (r_1 \ r_2) \cdot \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1^* \\ c_2^* \end{pmatrix} = r_1 \cdot (c_1^* + c_2^*) + 2 \cdot r_2 \cdot c_2^* = r_1 + 2 \cdot r_2 \cdot c_2^* \leq 1 = E(r^*, c^*), \quad \text{pois} \quad 0 \leq c_2^* \leq \frac{1}{2} \quad \text{e} \quad r_1 + r_2 = 1; \\ E(r^*, c) &= (1 \ 0) \cdot \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = c_1 + c_2 = 1 \geq E(r^*, c^*). \end{aligned}$$

Entretanto, não é razoável supor que o jogador das colunas escolha uma estratégia $c^* = \begin{pmatrix} c_1^* \\ c_2^* \end{pmatrix}$ com $0 < c_2^*$!

Dado um jogo representado por uma matriz $A = (a_{ij})$, dizemos que uma estratégia ótima r^* para o jogador das linhas **domina uma outra estratégia ótima** r para a toda estratégia pura e_j para o jogador das colunas $E(r, e_j) \leq E(r^*, e_j)$. Se $E(r, e_j) \leq E(r^*, e_j)$, para toda a estratégia ótima r e para toda e_j dizemos que r^* é dominante para o jogador das linhas. Se r^{**} é outra estratégia dominante, então $E(r^{**}, e_j) = E(r^*, e_j)$ para toda estratégia pura e_j para o jogador das colunas. Claro que uma estratégia dominante assegura para o jogador das linhas o maior *payoff* esperado: como $E(r, e_j) \leq E(r^*, e_j)$ para toda r e para toda e_j , então $E(r, c_1 \cdot e_1 + \dots + c_n \cdot e_n) \leq E(r^*, c_1 \cdot e_1 + \dots + c_n \cdot e_n)$, isto é, $E(r, c) \leq E(r^*, c)$ para todas as r ótimas e c , em particular, se (r^*, c^*) é uma sela, então $E(r^*, c^*) \leq E(r^*, c^*)$, como $E(r^*, c^*) \geq E(r^*, c^*)$, então se (r^*, c^*) é sela, então $E(r^*, c^*) = E(r^*, c^*)$. Na verdade, (r^*, c^*) é uma sela, pois $E(r, c^*) \leq E(r^*, c^*) = E(r^*, c^*) = E(r^*, c^*) \leq E(r^*, c) \leq E(r^*, c)$, isto é, $E(r, c^*) \leq E(r^*, c^*) \leq E(r^*, c)$. (r^*, c^*) .

Definimos da mesma maneira para o jogador das colunas: se c^* é ótima para o jogador das colunas, c^* é dominante se $E(e_i, c^*) \leq E(e_i, c)$, para toda a estratégia ótima c , para o jogador das colunas, e para toda estratégia pura e_i , para o jogador das linhas. No exemplo, a estratégia $c^* = e_1$ é dominante para o jogador das colunas, pois $E(e_1, e_1) = 1 \leq E(e_1, c) = c_1 + c_2 = 1$ e $E(e_2, e_1) = 0 \leq E(e_2, c) = 2 \cdot c_2$. Toda a estratégia dominante combinada com uma estratégia ótima para o oponente resulta num ponto de sela para os *payoffs* dos jogadores, mas nem toda estratégia numa sela é dominante. Por exemplo, $c^* = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ é tal que (e_1, c^*) é sela, mas c^* não é dominante para o jogador das colunas, pois $E(e_2, e_1) = 0 < E(e_2, c^*) = (0 \ 1) \cdot \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = 1$. Observe que todas as escolhas ótimas nesse jogo para o jogador das colunas são definidas por (e_i, c) com $c_2 \leq \frac{1}{2}$. Contudo, existe apenas uma dessas escolhas que é dominante: e_1 , pois $c_2 = 0$. Qual é o significado das estratégias dominantes? Elas preservam o jogador dos erros do oponente. No exemplo, se o jogador das colunas atribui alguma

probabilidade positiva à coluna dois, ele está se expondo ao risco de que um erro do jogador das linhas, ou seja, a escolha da linha dois, ocasione um prêmio de -2 . Se o jogador das linhas não cometer erros, o jogador das colunas recebe o prêmio -1 , independentemente da estratégia mista escolhida, mas se o jogador das linhas errar... Nesse jogo, entre as escolhas ótimas, ou seja, escolhas que constituem selas, existe a possibilidade de que um erro do oponente ocasione perdas.

Uma estratégia dominante r^* para o jogador das linhas é dita a **melhor estratégia** se existe uma estratégia pura $\bar{e}_j$ para o jogador das colunas, tal que $E(r, \bar{e}_j) < E(r^*, \bar{e}_j)$ para toda r ótima, $r \neq r^*$. Uma estratégia c^* é dita a **melhor estratégia para o jogador das colunas** se c^* é dominante e existe uma estratégia pura $\bar{e}_i$ para o jogador das linhas, tal que $E(\bar{e}_i, c^*) < E(\bar{e}_i, c)$ para toda a estratégia c ótima, $c \neq c^*$. Portanto, em jogos como o do exemplo, a melhor estratégia preserva o jogador de eventuais erros do oponente. No caso, e_1 é a melhor estratégia para o jogador das colunas, pois $E(e_2, e_1) = 0 < E(e_2, c) = 2 \cdot c_2$, se $c_2 > 0$.

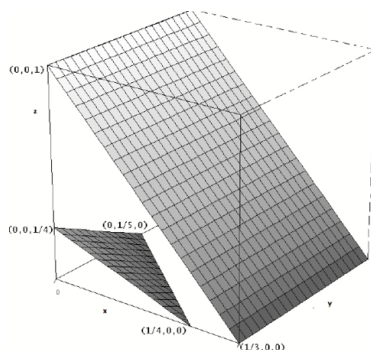
Exemplo 16: Retornando ao jogo $A = \begin{pmatrix} 4 & 5 & 4 \\ 3 & 0 & 1 \end{pmatrix}$, que é dado por Luce e Raiffa e que tem duas selas (e_1, e_1) e (e_1, e_3) . Para o jogador das colunas, a estratégia e_1 não é dominante, pois $E(e_2, e_1) = 3 > E(e_2, e_3) = 1$. As estratégias ótimas para o jogador das colunas são dadas por $c = \begin{pmatrix} 1-t \\ 0 \\ t \end{pmatrix}, 0 \leq t \leq 1$, e $E(e_1, e_3) = 4 \leq E(e_1, c) = 4 \cdot (1-t) + 5 \cdot 0 + 4 \cdot t = 4$ e $E(e_2, e_3) = 1 \leq E(e_2, c) = 3 \cdot (1-t) + 0 \cdot c_2 + t = 3 - 2 \cdot t$. Portanto, e_1 é dominante e como $E(e_2, e_3) = 1 < E(e_2, c) = 3 \cdot (1-t) + 0 \cdot c_2 + t = 3 - 2 \cdot t$ se $t < 1$, então e_3 é a melhor estratégia.

Se formulamos o problema do jogador das colunas, que é:

$$\begin{array}{ll} \max & x_1 + x_2 + x_3 \\ \text{sa} & 4 \cdot x_1 + 5 \cdot x_2 + 4 \cdot x_3 \leq 1 \\ & 3 \cdot x_1 + 0 \cdot x_2 + x_3 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, \end{array}$$

observamos que se $4 \cdot x_1 + 5 \cdot x_2 + 4 \cdot x_3 \leq 1$, então $3 \cdot x_1 + 0 \cdot x_2 + x_3 \leq 4 \cdot x_1 + 5 \cdot x_2 + 4 \cdot x_3$. Portanto, a única restrição do problema acima é $4 \cdot x_1 + 5 \cdot x_2 + 4 \cdot x_3 \leq 1$. Ora, o

valor máximo para a soma $x_1 + x_2 + x_3 = \frac{1}{v} = \frac{1}{4}$ está na aresta entre os vértices $(1, 0, 0)$ e $(0, 0, 1)$, então as estratégias $c = \begin{pmatrix} c_1 \\ 0 \\ c_3 \end{pmatrix}, c_1 + c_3 = 1, c_j \geq 0$, são min-max. Veja o desenho abaixo:



Considere o problema do jogador das linhas:

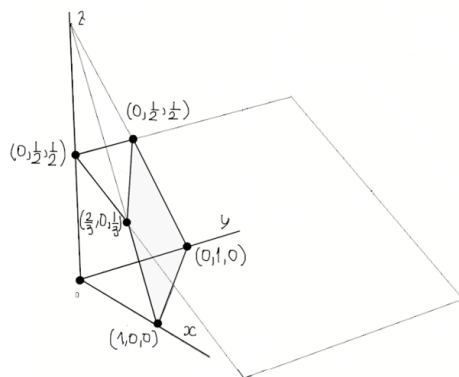
$$\begin{array}{ll} \min & y_1 + y_2 \\ \text{sa} & 4 \cdot y_1 + 3 \cdot y_2 \geq 1 \\ & 5 \cdot y_1 \geq 1 \\ & 4 \cdot y_1 + y_2 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

Os vértices da região acima estão em $(1/5, 1/5)$ e $(1/4, 0)$, portanto, para o jogador das linhas, existe apenas uma estratégia ótima que é $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Logo, os pares de estratégias $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $c = \begin{pmatrix} c_1 \\ 0 \\ c_3 \end{pmatrix}, c_1 + c_3 = 1, c_j \geq 0$, são os pontos de sela para o *payoff* esperado pelo jogador das linhas. Observe que, se o jogador das colunas proceder por escolher aleatoriamente entre a coluna um e a coluna dois, o nível de segurança permanece constante em quatro, ou seja, admitir as estratégias mistas não traz nenhum ganho para o jogador das colunas se o oponente empregar a estratégia max-min. Todavia, se temos a suspeita adicional que nosso oponente possa não empregar a estratégia max-min, então é vantajoso empregar $c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Exemplo 17: Não é difícil concluir que no jogo $A = \begin{pmatrix} 4 & 4 \\ 1 & 3 \\ 3 & 1 \end{pmatrix}$, a estratégia ótima para o jogador das linhas é $r^* = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ e que qualquer estratégia $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ é ótima para o jogador das colunas. Entretanto, nenhuma dessas estratégias é dominante. Nesse jogo não existe uma melhor estratégia para o jogador das colunas.

Exercício 10

i) Considere o jogo $A = \begin{pmatrix} 1 & 1 & 1 \\ \frac{1}{2} & 0 & 2 \end{pmatrix}$. Mostre que os pares de estratégias $r^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e $c^* = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$, $\frac{c_1}{2} + 2 \cdot c_3 \leq 1$ constituem uma sela para os *payoffs* esperados. Por exemplo, $c^* = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $c^* = \begin{pmatrix} 2/3 \\ 0 \\ 1/3 \end{pmatrix}$, $c^* = \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix}$ e $c^* = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ são estratégias ótimas para o jogador das colunas e são os vértices do quadrilátero sobre o plano $x_1 + x_2 + x_3 = 1$ que determinam as estratégias ótimas para o jogador das colunas.



No entanto, a melhor estratégia para o jogador das colunas é $c^* = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, pois $(0 \ 1) \cdot \begin{pmatrix} 1 & 1 & 1 \\ \frac{1}{2} & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 0$.

ii) Considere o jogo $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$. Observe que nesse jogo qualquer par $r^* = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$, $r_1 + r_2 = 1$, $c^* = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ é uma sela para os *payoffs* esperados e que $r^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ é a

melhor estratégia para o jogador das linhas, protegendo-o de prováveis erros do oponente.

iii) Considere o jogo $A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}$. Observe que nesse jogo o jogador das colunas jamais escolherá a coluna dois. Portanto, esse jogo pode ser reduzido ao jogo $A = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}$. Mostre que $r^* = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}$, $c^* = \begin{pmatrix} 2/3 \\ 0 \\ 1/3 \end{pmatrix}$ é uma sela para os *payoffs* esperados e que não existem estratégias ótimas puras para esses jogadores.

Alguns autores acham insatisfatória a justificativa das probabilidades pela necessidade do segredo, como Luce e Raiffa no livro “*Games and Decisions, Introduction and Critical Survey*” (1957):

O argumento mais comum em defesa de estratégias mistas é a observação de que elas impossibilitam nossos oponentes de conhecer a estratégia pura que usaremos. Isso, afirma-se, é importante, pois esse conhecimento pode ser explorado [pelo oponente]. Embora admitindo que esse seja frequentemente um ponto pertinente, alguns autores acham que certamente não pode ser uma defesa completa, pois estratégias mistas são apropriadas mesmo quando sabemos que o outro jogador não está de maneira alguma preocupado com o comportamento do oponente. Isso pode ocorrer se o outro jogador não estiver ciente das recompensas do jogo ou em jogos contra a natureza [ou seja, contra um oponente inconsciente].

E, perguntam-se:

Mas, se a defesa de estratégias mistas não deveria se limitar apenas a um argumento de sigilo, em que deveria basear-se? A resposta é que estratégias mistas nos impedem sermos vítimas da nossa fragilidade humana. Não é diferente da pessoa que quer fazer dieta. Ele anuncia sua intenção ou aceita uma aposta de que não interromperá sua dieta, para que mais tarde não seja livre para mudar de ideia e otimize suas ações de acordo com seus gostos naquele momento – por exemplo, comer um *sundae* de sorvete.³⁵

Mas, os autores reconhecem que nem todos aceitam esse argumento e limitam-se à justificativa do segredo.³⁶

Luce e Raiffa insistem que **não existe prioridade entre estratégias puras e mistas:**

Embora não haja conflito de interesses entre pares de equilíbrio em estratégias mistas e entre pares de equilíbrio em estratégias puras, quando existem, pode-se temer que haja conflito entre as duas classes de pares de equilíbrio. Em outras palavras, pode-se temer que a teoria das estratégias mistas e a teoria das estratégias puras sejam diferentes para os jogos com pares de equilíbrio em estratégias puras, que o nível ótimo de segurança usando estratégias mistas seja diferente daquele usando estratégias puras [veja a observação logo abaixo]. Mostra-se que esse não é o caso. Além disso, os pares de equilíbrio em estratégias puras também são pares de equilíbrio de estratégia mista; portanto, com relação aos melhores níveis de segurança, nenhuma complicação resulta da introdução de estratégias mistas em jogos com pares de equilíbrio em estratégias puras. No entanto, continua sendo verdade que, embora um jogador tenha uma estratégia de equilíbrio puro e a reconheça, ele pode ter motivos para jogar de forma diferente se achar que seu oponente possa jogar uma estratégia que não é de equilíbrio [é o caso do exercício 1, veja Exemplo 15]. O jogador pode afastar-se de uma estratégia de equilíbrio para explorar a suposta ignorância de um oponente, além disto pode ser sensato usar uma estratégia mista como uma proteção contra situações extremamente desfavoráveis e contra a possibilidade de que o oponente tenha mais conhecimento sobre o comportamento do que o previsto.³⁷

A expressão acima “Mostra-se que esse não é o caso” se refere ao fato de que o valor do jogo v é independente das selas. Digamos que (r^*, c^*) e (r'', c'') são selas para E , então $E(r'', c'') \leq E(r'', c^*) \leq E(r^*, c^*) \leq E(r^*, c'') \leq E(r'', c'')$, logo $E(r'', c'') \leq E(r^*, c^*) \leq E(r'', c'')$, portanto $E(r^*, c^*) = E(r'', c'') = v$.

Exercício 11

Usando a definição de sela $E(r, c^*) \leq E(r^*, c^*) \leq E(r^*, c)$, procure perceber que $E(r'', c'') \leq E(r'', c^*) \leq E(r^*, c^*) \leq E(r^*, c'') \leq E(r'', c'')$.

Von Neumann elaborou um critério de escolha dentre as opções possíveis mantendo o princípio maxmin-minmax em relação ao *payoff* esperado. Mostrar que isso é possível é o célebre resultado de von Neumann. A ideia é que a intuição do princípio maxmin-minmax deve ser mantida, isto é, se um jogador escolher a opção correta, não pode receber menos que um valor determinado, chamado valor do jogo.

A maneira como von Neumann imagina jogos promove uma profunda alteração sobre como concebemos jogos em geral. Quando

pensamos no comportamento dos jogadores, pensamos que estes sempre buscam a vitória, ou ainda, o máximo de ganho possível, como disse Borel. Von Neumann observa que isso pode não ser factível e altera a percepção sobre os objetivos dos jogadores: o que podem buscar consistentemente é minimizar suas perdas. Para von Neumann é claro que sua proposta tem um cunho normativo. Ele não pretende determinar como jogadores reais se comportam em jogos reais, mas recomendar uma regra aos jogadores quando as maneiras de jogar não são óbvias.

Von Neumann é um pessimista. Os jogadores de von Neumann são cautelosos, preocupados em minimizar perdas e não em assegurar vitórias, ainda que ele dissesse que estamos apresentando uma falsa oposição. Von Neumann viveu em tempos sombrios:

para o tipo de explosão que o homem será capaz de inventar em 1980, o mundo é perigosamente pequeno, sua unidade política é perigosamente instável [...] o fato concreto é que as dificuldades são devidas a uma evolução [tecnológica], que embora, útil e construtiva, também é perigosa. Podemos produzir os ajustes necessários com a velocidade necessária? Podemos especificar apenas qualidades humanas necessárias: paciência, flexibilidade, inteligência.³⁸

A Teoria dos Jogos é uma teoria normativa:

O propósito de uma teoria da tomada de decisão, ou especificamente da teoria dos jogos, é aconselhar uma pessoa como se comportar escolhendo de maneira ótima no conjunto das estratégias disponíveis para ela nas situações sujeitas à teoria. Se ela decidir se desviar conscientemente do curso indicado, se ela tem outro objetivo ou opõem-se aos meios a serem empregados (por razões morais e outras). É então uma questão de terminologia se ela é ainda considerada um ator racional. A teoria, de qualquer forma, pode levar tais desvios em consideração. Claramente, algumas estratégias tecnicamente disponíveis podem ser inadmissíveis por aspectos legais, morais e outros. Em alguns casos, essas questões não surgem: o xadrez é jogado da mesma maneira, quer os oponentes sejam ricos ou pobres. Católicos, maometanos, comunistas ou capitalistas. Entretanto, negócios ou negócios políticos são afetados por tais circunstâncias. O conselho pode ser dado com ou sem restrições que envolvam moral ou religião.³⁹

Todavia, Morgenstern também defende que a teoria é descritiva em algumas situações:

A teoria também é descritiva, como precisa ser para ser usada como modelo. Pode-se argumentar que a teoria não pode descrever eventos passados, pois antes de ser criada, os indivíduos não poderiam ter seguido a estratégia ótima que somente a teoria poderia descobrir! A resposta é que, para algumas situações, o indivíduo pode encontrá-la por tentativa e erro e uma tradição de conhecimento empírico pode se desenvolver após repetidas tentativas.⁴⁰

Não é uma argumentação muito convincente, mas, por exemplo, é empregada por Nash em sua tese de doutorado (1950) uma seção intitulada “*Motivation and Interpretation*”, ainda que ele nunca mais tenha se referido a ela.

Exercício 12

Considere o jogo:

$$A = \begin{pmatrix} 2 & 1 \\ 2 & 6 \\ 1 & 2 \end{pmatrix}.$$

i) Observe que $r^* = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $c^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $v^* = 2$.

ii) O problema do jogador das colunas é:

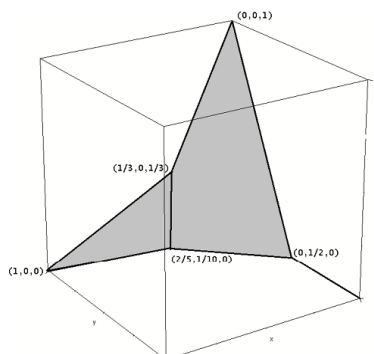
$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{sa} & 2 \cdot x_1 + x_2 \leq 1 \\ & 2 \cdot x_1 + 6 \cdot x_2 \leq 1 \\ & x_1 + 2 \cdot x_2 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Observando as restrições acima, por que a única restrição efetiva é $2 \cdot x_1 + 6 \cdot x_2 \leq 1$? Por que é fácil observar que a estratégia do jogador das colunas é $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ através do problema acima?

iii) O problema do jogador das linhas é:

$$\begin{array}{ll} \min & y_1 + y_2 + y_3 \\ \text{sa} & 2 \cdot y_1 + 2 \cdot y_2 + y_3 \geq 1 \\ & y_1 + 6 \cdot y_2 + 2 \cdot y_3 \geq 1 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

Abaixo está desenhado o simplexo definido pelas restrições. Qual é a solução do problema acima?



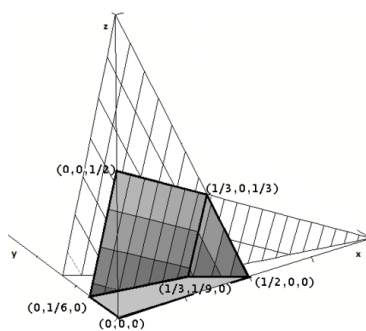
iv) Observe que as soluções do problema acima estão na aresta determinada pelo segmento de reta que une os pontos $(0,1/2,0)$ e $(2/5,1/10,0)$.

v) Observe que $r = \begin{pmatrix} r_1 \\ r_2 \\ 0 \end{pmatrix}$, $r_1 + r_2 = 1$, $0 \leq r_1 \leq \frac{4}{5}$ e $c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ são selas de E .

Exercício 13

O problema do jogador das colunas é:

$$\begin{aligned} \max \quad & x_1 + x_2 + x_3 \\ \text{sa} \quad & 2 \cdot x_1 + 3 \cdot x_2 + x_3 \leq 1 \\ & x_1 + 6 \cdot x_2 + 2 \cdot x_3 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{aligned}$$



i) Resolva o problema acima e encontre $c^* = \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix}$.

ii) O problema do jogador das linhas é:

$$\begin{array}{ll}
\min & y_1 + y_2 \\
\text{sa} & 2 \cdot y_1 + y_2 \geq 1 \\
& 3 \cdot y_1 + 6 \cdot y_2 \geq 1 \\
& y_1 + 2 \cdot y_2 \geq 1 \\
& y_1 \geq 0, y_2 \geq 0.
\end{array}$$

iii) Resolva o problema acima e encontre $r^* = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$.

iv) Observe que (r^*, c^*) é a única sela de E .

Exercício 14

Considere o jogo $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \end{pmatrix}$.

i) Verifique que o *payoff* esperado pelo jogador das linhas é $E(r, c) = r' \cdot A \cdot c = 1 + r_2 \cdot (c_2 + c_3)$.

ii) A estratégia minmax por parte do jogador das colunas é obtida da seguinte maneira: fixada uma estratégia $c = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$, o jogador das colunas busca a estratégia por parte do jogador das linhas que mais o prejudica (nível de segurança), ou seja,

$$\begin{array}{ll}
\max & 1 + r_2 \cdot (c_2 + c_3) \\
\text{sa} & r = r_1 + r_2 = 1 \\
& r_1 \geq 0, r_2 \geq 0.
\end{array}$$

iii) Observe que a solução está em $r_2 = 1$. Na sequência, o jogador das colunas minimiza $1 + c_2 + c_3$. Qual é a estratégia ótima para o jogador das colunas?

iv) A estratégia maxmin por parte do jogador das colunas é obtida da seguinte maneira: fixada uma estratégia $r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$, o jogador das linhas busca a estratégia por parte do jogador das colunas que mais o prejudica, ou seja,

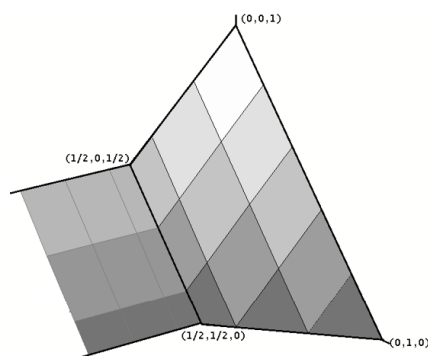
$$\begin{array}{ll}
\min & 1 + r_2 \cdot (c_2 + c_3) \\
\text{sa} & c = c_1 + c_2 + c_3 = 1 \\
& c_1 \geq 0, c_2 \geq 0, c_3 \geq 0.
\end{array}$$

iv) Observe que a solução está em $c_1 = 1$ e $c_2 + c_3 = 0$. Por que? Isto resulta num *payoff* igual a um que independe da escolha de $r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$. Quais são as estratégias ótimas para o jogador das linhas?

v) Mostre que $r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \in \Delta_2$ e $c = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ definem selas para E . Nesse exemplo, todas as selas estão satisfazendo as soluções maxmin e minmax.

Exercício 15

Olhando o simplexo abaixo determinado pelas restrições do jogador das linhas, obtenha a matriz do jogo.



Resposta: $A = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}$.

Exercício 16

Dada a matriz $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, o *payoff* esperado pelo jogador das linhas, que escolhe $p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \in \Delta_2$, é dado por:

$$K(p, q) = p' \cdot A \cdot q = (p_1 \quad p_2) \cdot \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ q_2 \end{pmatrix},$$

em que $q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \in \Delta_2$ é escolhido pelo jogador das colunas.

i) Mostre que:

$$K = (a_{11} - a_{12} - a_{21} + a_{22}) p_1 \cdot q_1 + (a_{12} - a_{22}) \cdot p_1 + (a_{21} - a_{22}) \cdot q_1 + a_{22};$$

ii) Pelo princípio maxmin, fixado um $p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$, então o *payoff* é minimizado. Caso $0 < q_1 < 1$, então K é minimizado em $\frac{\partial K}{\partial q_1} = 0$. Obtenha p_1^* e p_2^* ótimos;

$$p_1^* = \frac{a_{22} - a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}; \quad p_2^* = \frac{a_{11} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}}.$$

iii) Esse método não funciona se $0 = q_1^*$ ou $q_1^* = 1$. Por exemplo, se $A = \begin{pmatrix} 0 & -1 \\ 5 & 1 \end{pmatrix}$, é evidente que $p^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e $q^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ são as soluções do jogo. Verifique isso. Mas, nosso método produz $p_1 = \frac{4}{3} > 1$, o que é um absurdo. Verifique isso. Por que o método não funciona se as soluções são estratégias puras?

iv) Caso $0 < p_1^* < 1$, então K é maximizado em $\frac{\partial K}{\partial p_1}$. Obtenha q_1^* e q_2^* ótimos:

$$q_1^* = \frac{a_{22} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}}; \quad q_2^* = \frac{a_{11} - a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}.$$

v) Usando a expressão

$$K = (a_{11} - a_{12} - a_{21} + a_{22}) p_1 \cdot q_1 + (a_{12} - a_{22}) \cdot p_1 + (a_{21} - a_{22}) \cdot q_1 + a_{22},$$

verifique que o valor do jogo é dado por

$$v = \frac{a_{22} \cdot a_{11} - a_{21} \cdot a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}} = \frac{\det A}{a_{11} - a_{12} - a_{21} + a_{22}}.$$

Exercício 17

No parágrafo 18.4.4 de “*Theory of Games and Economic Behavior*”, von Neumann e Morgenstern apresentam o seguinte exemplo:

Consideremos um jogo sugerido por um episódio do livro “As Aventuras de Sherlock Holmes”. Sherlock Holmes deseja seguir de Londres para Dover e daí para o continente visando escapar do professor Moriarty [que é professor de Matemática...] que o persegue. Tendo embarcado no trem, ele vê, quando o trem parte, a silhueta do professor Moriarty na plataforma. Sherlock Holmes assume como certo que viu Moriarty e que está totalmente justificado em presumir que seu adversário também o viu e que pegará um trem rápido visando ultrapassá-lo. Sherlock Holmes se depara com a alternativa de ir para Dover ou deixar o trem em Canterbury, a única estação intermediária. Seu adversário cuja inteligência é considerada adequada para visualizar essas possibilidades tem as mesmas escolhas. Ambos os oponentes devem escolher o local de desembarque,

ignorando a decisão do outro. Se, como resultado dessas decisões, eles se encontrarem na mesma plataforma, Sherlock Holmes supõe que será certamente morto por Moriarty. Se Sherlock Holmes chegar a Dover ileso, ele escapará. Quais são as estratégias adequadas particularmente para Sherlock Holmes?⁴¹ (tradução do autor)

Von Neumann e Morgenstern propõem a seguinte matriz de *payoffs* para o jogador das linhas (Moriarty):

	Sherlock desembarca em Dover	Sherlock desembarca em Caterbury
Moriarty desembarca em Dover	100 Sherlock morre	0 Sherlock escapa
Moriarty desembarca em Caterbury	-50 Sherlock escapa	100 Sherlock morre

O resultado que von Neumann e Morgenstern obtém é o seguinte: $r = \left(\frac{3}{5}\right)$ e $c = \left(\frac{2}{5}\right)$. Verifique isso. Qual é o valor do jogo? A observação de von Neumann e Morgenstern: “Moriarty deve ir para Dover com uma probabilidade de 60%, enquanto Sherlock Holmes deve parar na estação intermediária com uma probabilidade de 60%, ficando os 40% restantes em cada caso para a outra alternativa.”⁴²; numa nota de pé de página, os autores observam:

Desculpavelmente, a narrativa de Conan Doyle [o criador de Sherlock e Moriarty] **desconsidera as estratégias mistas e narra os acontecimentos reais.** De acordo com estes, Sherlock Holmes sai na estação intermediária e vê triunfante o trem especial de Moriarty indo para Dover. A solução de Conan Doyle é a melhor possível sob suas limitações (estratégias puras), na medida em que ele atribui a cada oponente o percurso que achamos ser o mais provável (ou seja, ele substitui 60% de probabilidade por certeza). No entanto, é um pouco enganador que esse procedimento leve à vitória completa de Sherlock Holmes, na medida em que, como vimos acima, as probabilidades (ou seja, o valor do jogo) são definitivamente a favor

de Moriarty. (Nosso resultado indica que Sherlock tem 48% de chance de estar morto quando pula do trem que veio da Victoria Station [ponto de partida em Londres]).⁴³

Verifique essas afirmações. A probabilidade de Sherlock morrer é $P = \frac{3}{5} \cdot 100 \cdot \frac{2}{5} + \frac{2}{5} \cdot 100 \cdot \frac{3}{5}$.

Exercício 18

Num artigo clássico “*An Experimental Analysis of Some Variables of Minimax Theory*”, resultado de sua tese de doutorado, Richard Brayer investiga se jogadores em jogos de dois jogadores de soma-zero efetivamente escolhem a estratégia maxmin-minmax em estratégias puras como Borel, von Neumann e Morgenstern supõem:

A teoria dos jogos não é descritiva. Esta teoria não prescreve estratégias para alcançar objetivos específicos dentro de um conjunto específico de regras, não faz declarações sobre as estratégias que as pessoas realmente usam para jogar, se suas estratégias estão em conformidade com as sugeridas pelo modelo teórico do jogo ou quais fatores, além das regras do jogo, podem afetar a estratégia de um jogador. [...] É importante enfatizar que a teoria dos jogos é uma teoria normativa. Dadas certas situações de jogo, sugere métodos de jogo para atingir certos objetivos prudentes. **A teoria dos jogos não é descritiva.**⁴⁴ (grifo e tradução do autor)

Brayer apresenta um exemplo dos jogos aos quais foram submetidos 100 estudantes homens da University of Massachusetts:

	e_1	e_2	e_3
$s_1 \rightarrow$	11	-7	8
$s_2 \rightarrow$	1	1	2
$s_3 \rightarrow$	-10	-7	21

As linhas são as estratégias disponíveis para os estudantes e as colunas, para o experimentador. Qual é a sela em estratégias puras desse jogo? Qual é o valor do jogo? As conclusões do experimento foram:

Em geral, os resultados deste estudo sustentam a controvérsia de que a estratégia escolhida por alguém depende não apenas da natureza teórica do jogo, mas também das

condições sob as quais é jogado. Sob certas condições, os Ss [os jogadores submetidos ao teste e que escolhem linhas] estavam de acordo com a jogada sugerida para um jogo de duas pessoas de soma-zero e estritamente determinado [ou seja, uma sela em estratégias puras]. Sob outras condições, eles se ajustavam bastante menos. Contra um oponente racional, que joga minimax, os Ss aprenderam a jogar a solução sugerida pelo ponto de sela. Os dados indicariam, porém, que os Ss não aprenderam a lógica minimax do jogo, mas foram apenas respondendo à estratégia de seu oponente, ou seja, eles estavam jogando uma contra-estratégia. Embora contra um oponente aleatório, os Ss não jogassem a alternativa minimax (alta média) de forma consistente, eles escolheram essa alternativa mais do que qualquer uma das outras. Os dados indicam ainda que os Ss que foram informados de que seu oponente estava jogando aleatoriamente tendem a se ajustar com mais frequência à estratégia minimax. Aqueles Ss que disseram que seu oponente era racional, embora ele realmente jogasse aleatoriamente, tendiam a procurar estratégias no jogo de seu oponente. Esse grupo se ajustou menos à estratégia sugerida, pois se alguém assume que seu oponente é racional, a alternativa minimax é a solução de sela para o oponente. Teoricamente, o valor do jogo não deveria afetar as escolhas de um jogador. Os dados deste estudo indicam uma tendência por parte dos Ss para escolherem a alternativa do ponto de sela para os jogos de maior valor.⁴⁵ (tradução do autor)

4. A polêmica

Em 1953, o matemático Maurice Fréchet (1878-1973) publicou uma tradução para o inglês dos três artigos sobre teoria dos jogos do conterrâneo Émile Borel (1871-1956). Os artigos originais foram publicados em 1921⁴⁶, em 1924⁴⁷ e em 1927⁴⁸. Fréchet⁴⁹ apresenta os artigos de Borel com um comentário no qual minimiza o trabalho de von Neumann em Teoria dos Jogos. Fréchet observou que o teorema minimax seria uma consequência dos trabalhos de Borel, ainda que o mesmo não tenha obtido a demonstração desse resultado e até mesmo não tenha acreditado na possibilidade de obtê-la. De qualquer maneira, o anteviu e, além disso, o resultado foi obtido através da análise convexa por Jean Ville em 1937, um orientando de Borel.

Von Neumann⁵⁰ rebateu as críticas de Fréchet observando que obter determinado resultado como consequência de trabalhos

anteriores não tira a prioridade do autor (von Neumann), pois apenas posteriormente se pode observar esse fato. Von Neumann ainda afirmou que desconhecia o trabalho de Borel em 1928 e defendeu sua prioridade no teorema minimax.

5. O guia von Neumann para Economia

i) O artigo *“Zur Theorie der Gesellschaftsspiele”* foi publicado em 1928, traduzido para o inglês por Sonya Bargmann com o título *“On the Theory of Games of Strategy”* e publicado em 1959, abrindo o volume IV de *“Contributions to The Theory of Games”*, nº 40, dos *“Annals of Mathematic Studies”*. Von Neumann falecera dois anos antes, em 1957. Os resultados principais desse artigo são:

- 1) definição de jogos na forma extensiva e normal;
- 2) prova da existência de equilíbrio em jogos de soma-zero com dois jogadores (teorema minimax);
- 3) introdução da noção de funções quase-côncavas e quase-convexas;
- 4) discussão de jogos de soma-zero com mais de dois jogadores.

ii) O artigo *“Über ein Ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes”* foi publicado em 1937, mas von Neumann afirma que o trabalho foi apresentado pela primeira vez em 1932, num seminário na Universidade de Princeton. Esse artigo foi traduzido para o inglês e publicado na *“The Review of Economic Studies”*, em 1945-1946, sob o título *“A Model of General Economic Equilibrium”*⁵¹. Os resultados principais desse artigo são:

- 1) introdução de inequações em lugar de equações;

- 2) prova de existência de equilíbrio via Teorema do Ponto Fixo de Brouwer, o que propiciou o trabalho posterior de Nash e as demonstrações de equilíbrio geral de Arrow, Debreu e MacKenzie;
- 3) sugestão de uma nova prova do teorema minimax via ponto fixo.

Talvez, o mais importante trabalho para a Economia Matemática no século 20 tanto para a teoria ortodoxa quanto para as versões matemáticas de sraffianos e marxistas.

iii) No *"Theory of Games and Economic Behavior"*, publicado em 1944, destacamos:

- 1) nova prova do teorema minimax;
- 2) aprofundamento das seções finais do artigo de 1928 para jogos com mais de dois jogadores; desenvolvimento dos jogos cooperativos;
- 3) o conceito de utilidade esperada.

iv) No artigo *"Discussion of a Maximum Problem"*, de 1947, publicado no volume 6 dos *"Collected Works"*, destacamos:

- 1) a introdução do lema de Farkas (1901), possibilitando uma nova prova do teorema minimax;
- 2) fundamentação teórica da programação linear;
- 3) entendimento da dualidade em problemas de programação linear.

v) A quarta demonstração do Teorema Minimax apareceu no artigo *"Solutions of Games by Differential Equations"*, em coautoria com G. W. Brown, publicado em 1950 no n° 24 do *"Annals of Mathematic Studies"*, volume I do *"Contributions To The Theory of Games"*.

vi) “A Certain Zero-Sum Two-Person Game Equivalent to Optimal Assignment Problem”, publicado em 1953 no n° 28 do “*Annals of Mathematic Studies*”, volume II do “*Contributions To The Theory of Games*”.

vii) “A Numerical Method to Determine Optimum Strategy”, publicado em 1954 no “*Naval Research Logistics Quartely*”, volume 1, n° 2.

No volume 6 dos “*Collected Works*”⁵² sobre jogos, ainda estão publicados os seguintes artigos: “*Numerical Method for Determination of Value and Best Strategies of 0-Sum 2-Person Game with Large Number of Strategies*” (1946); “*Two variants of Poker*” (1953); “*Symmetric solutions of some general N person games*” (sem data).

Segundo a bibliografia elaborada por Stanislaw Ulam⁵³, no ano de 1928, von Neumann publicou ainda outros nove trabalhos, todos relacionados a Mecânica Quântica.

¹ DANTZIG, G. “A Proof of the Equivalence of the Programming Problem and the Game Problem”. [Chapter XX]. In: KOOPMANS, T. *Activities Analysis of Production and Allocation*. New York: John Wiley & Sons, 1951, p. 33.

² SCHWALBE, U.; WALKER, P. “Zermelo and Eearly History of Game Theory”. **Games and Economic Behavior**, v. 34, n° 1 (2001), pp. 123-137. [No Apêndice, os autores traduzem para o inglês o artigo original de Zermelo].

³ BOREL, E. “The Theory of Play and Integral Equations with Symmetric Kernels”. **Econometrica**, v. 21, n°1 (1953), pp. 97-100.

BOREL, E. “On the Games that involve Chances and the Skill of the Players”. **Econometrica**, v. 21, n° 1 (1953), pp. 101-115.

BOREL, E. “On Systems of Linear Forms of Skew Symmetric Determinant and General Theory of Play”. **Econometrica**, v. 21, n° 1 (1953), pp. 116-116.

FRÉCHET, M. “Emile Borel, Initiator of the Theory of Psychological Games and Its Application”. **Econometrica**, v. 21, n° 1 (1953), pp. 95-96.

- NEUMANN, J.; FRÉCHET, M. "Communication on the Borel Notes". **Econometrica**, v. 21, nº 1 (1953), pp. 124-127.
- ⁴ NEUMANN, J. "Zur Theorie der Gesellschaftsspiele". **Mathematische Annalen**, v. 100 (1928), pp. 295-320.
- ⁵ HALMOS, P. "The Legend of von Neumann". **The American Mathematical Monthly**, v. 80, nº 4 (1973), pp. 382-394.
- ⁶ SMITH, P. **Os Homens do Fim do Mundo: o Verdadeiro Dr. Fantástico e o Sonho da Arma Total**. São Paulo: Companhia das Letras, 2007, p. 394.
- ⁷ NEUMANN, J.; MORGENSTERN, O. **Theory of Games and Economic Behavior**. 7ª ed. Princeton: Princeton University Press, 1953.
- ⁸ MORGENSTERN, O. "Game Theory". In: WIENER, P. (Ed.) **Dictionary of the History of Ideas, Studies of Select Pivotal Ideas**, v. II. New York: Charles Scribner's Sons, 1973, pp. 263-275.
- ⁹ NASH, J. "Equilibrium Points in N-Person Games". **Proceedings of the National Academy of Sciences of the United States of America**, v. 36, nº 1 (1950), pp. 48-49.
- NASH, J. "The Bargaining Problem". **Econometrica**, v. 18, nº 2 (1950), pp. 155-162.
- NASH, J. "Non-Cooperative Games". **The Annals of Mathematics**, v. 54, nº 2 (1951), pp. 286-295.
- NASH, J. "Two-Person Cooperative Games". **Econometrica**, v. 21, nº 1 (1953), pp. 128-140.
- ¹⁰ MCKINSEY, J.C. C. **Introduction to the Theory of Games**. California: RAND Corporation, 1952.
- ¹¹ SMITH, J. M. **Evolution and the Theory of Games**. Londres: Cambridge University Press, 1982.
- ¹² KREPS, D. M. **A Course in Microeconomic Theory**. New York: Princeton University Press, 1990.
- ¹³ BAIRD, D.; GERTNER, R.; PICKER, R. **Game Theory and the Law**. Cambridge: Harvard University Press, 1994.
- ¹⁴ NEUMANN, J. "Calcul des Probabilités – Sur la théorie des Jeux". [Note de M. J. v. Neumann, présentée par M. Émile Borel]. **Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences** [Relatórios Semanais das Sessões da Academia de Ciências], v. 186, nº 25 (1928), pp. 1689-1691.
- ¹⁵ BOREL, E. "The Theory of Play and Integral Equations with Skew Symmetric Kernels". **Econometrica**, v. 21, nº 1 (1953a), pp. 97-100.
- BOREL, E. "On Games that Involve Chance and the Skill of the Players". **Econometrica**, v. 21, nº 1 (1953b), pp. 101-115.
- BOREL, E. "On Systems of Linear Forms of Skew Symmetric Determinant and the General Theory of Play". **Econometrica**, v. 21, nº 1 (1953c), pp. 116-117.

- ¹⁶ NEUMANN, J. “Zur Theorie der Gesellschaftsspiele”. **Mathematische Annalen**, v. 100, n° 1 (1928), pp. 295–320. Esse artigo foi traduzido para o inglês por Sonya Bargmann com o título “*On the Theory of Games of Strategy*” e abre o volume IV do “*Contributions to The Theory of Games*”, n° 40 dos “*Annals of Mathematic Studies*” de 1959.
- ¹⁷ NEUMANN, J., MORGENSTERN, O. **Theory of Games and Economic Behavior**. 7ª ed. Princeton: Princeton University Press, 1953.
- ¹⁸ LUCE, R. D., RAIFFA, H. **Games and Decisions, Introduction and Critical Survey**. New York: Dover Publications, INC, 1989, p. 66.
- ¹⁹ NEUMANN, J. “Calcul des Probabilités – Sur la théorie des Jeux”. [Note de M. J. v. Neumann, présentée par M. Émile Borel]. **Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences** [Relatórios Semanais das Sessões da Academia de Ciências], v. 186, n° 25 (1928), pp. 1689-1691.
- ²⁰ NEUMANN, J.; MORGENSTERN, O. **Theory of Games and Economic Behavior**. 7ª ed. Princeton: Princeton University Press, 1953 pp. 143-144.
- ²¹ NEUMANN, J.; MORGENSTERN, O. **Theory of Games and Economic Behavior**. 7ª ed. Princeton: Princeton University Press, 1953 pp. 145-146.
- ²² MCKINSEY, J.C. C. **Introduction to the Theory of Games**. California: RAND Corporation, 1952.
- ²³ http://mphpa.org/classic/VET_ARCHIVES/More/16-04_haywood.htm
- ²⁴ HAYWOOD JR., O. G. “Military Decision and Game Theory”. **Journal of the Operations Research Society of America**, v. 2, n° 4 (1954), p. 365.
- ²⁵ Ibidem, p. 372.
- ²⁶ Ibidem, p. 373.
- ²⁷ Ibid., p. 374
- ²⁸ NEUMANN, J. “On the Theory of Games of Strategy”. In: TUCKER, A.; LUCE, R. (Ed.). **Contributions to the Theory of Games**, v. 4. Annals of Mathematics Studies. 40ª ed. Princeton: Princeton University Press, 1959, p. 22.
- ²⁹ LUCE, R. D.; RAIFFA, H. **Games and Decisions, Introduction and Critical Survey**. New York: Dover Publications, INC, 1989, p. 76.
- ³⁰ HAYWOOD JR., O. G. “Military Decision and Game Theory”. **Journal of the Operations Research Society of America**, v. 2, n° 4 (1954), p. 377.
- ³¹ Ibidem, p. 377.
- ³² Ibidem, p. 378.
- ³³ HAYWOOD JR., O. G. “Military Decision and Game Theory”. **Journal of the Operations Research Society of America**, v. 2, n° 4 (1954).
- ³⁴ BRONOWSKI, J. **A Escalada do Homem**. 3ª ed. São Paulo: Livraria Martins Fontes Editora Ltda, 1972, p. 257.

- ³⁵ LUCE, R. D.; RAIFFA, H. **Games and Decisions, Introduction and Critical Survey**. New York: Dover Publications, INC, 1989, p. 75.
- ³⁶ Ibidem, p. 76.
- ³⁷ Ibidem, p. 73.
- ³⁸ NEUMANN, J. "Can we Survive Tecnology?". **Collected Works**. Vol. VI. Londres: Pergamon Press LTD, 1963, pp. 504-519.
- ³⁹ MORGENSTERN, O. "Game Theory". In: WIENER, P. (Ed.) **Dictionary of the History of Ideas, Studies of Select Pivotal Ideas**. Vol. II. New York: Charles Scribner's Sons, 1973, p. 268.
- ⁴⁰ Ibidem, p. 268.
- ⁴¹ NEUMANN, J., MORGENSTERN, O. **Theory of Games and Economic Behavior**. 7^a ed. Princeton: Princeton University Press, 1953, p. 176.
- ⁴² NEUMANN, J., MORGENSTERN, O. **Theory of Games and Economic Behavior**. 7^a ed. Princeton: Princeton University Press, 1953.
- ⁴³ Ibidem, p. 178.
- ⁴⁴ BRAYER, R. "An Experimental Analysis of Some Variables of Minimax Theory". **Behavioral Sciences**, v. 9, n^o 1 (1964), p. 33.
- ⁴⁵ Ibidem, p. 43.
- ⁴⁶ BOREL, E. "The Theory of Play and Integral Equations with Symmetric Kernels". **Econometrica**, v. 21, n^o 1 (1953), pp. 97-100.
- ⁴⁷ BOREL, E. "On the Games that involve Chances and the Skill of the Players". **Econometrica**, v. 21, n^o 1 (1953), pp. 101-115.
- ⁴⁸ BOREL, E. "On Systems of Linear Forms of Skew Symmetric Determinant and General Theory of Play". **Econometrica**, v. 21, n^o 1 (1953), p. 116.
- ⁴⁹ FRÉCHET, M. "Emile Borel, Initiator of the Theory of Psychological Games and Its Application." **Econometrica**, v. 21, n^o 1 (1953), pp. 95-96.
- ⁵⁰ NEUMANN, J.; FRÉCHET, M. "Communication on the Borel Notes". **Econometrica**, v. 21, n^o 1 (1953), pp. 124-127.
- ⁵¹ NEUMANN, J. "A Model of General Economic Equilibrium". *The Review of Economic Studies*, v. 13, n^o 1 (1945-1946), pp. 1-9.
- ⁵² Neumann, J. **Collected Works**. Volume VI. Londres: Pergamon Press LTD, 1963.
- ⁵³ ULAM, S. "Tribute to John von Neumann". **Bulletin of the American Mathematical Society**, v. 64, n^o 3 (1958), pp. 1-49.

Lição TRÊS

O Método Simplex

Sumário

Lição 3 – O Método Simplex

- 1 Introdução
2. O Método Simplex e a linha-redução de sistemas lineares
3. Obtendo os vértices de polígonos
4. Encontrando os pontos ótimos em programação linear
5. A dualidade
6. Usando o Método Simplex para resolver o problema dual diretamente
7. O exemplo de Beale
8. Dantzig e o Método Simplex
9. George Bernard Dantzig (1914-2005)
10. A Pesquisa Operacional (OR)

Lição 3

O Método Simplex

1 Introdução

O Método Simplex aparece no capítulo XXI do livro “*Activity Analysis of Production and Allocation*”¹, no artigo “*Maximization of a Linear Function of Variables Subject to Linear Inequalities*”, de George Dantzig (1914-2005). Em nota, no artigo, Dantzig registra que o Método Simplex resultou de suas discussões com Leonid Hurwicz, Tjalling C. Koopmans e Marshall K. Wood em 1947. Também von Neumann participou dessas discussões e Dantzig relembra que, em 1947, visitou-o em Princeton para examinar os problemas relacionados à programação linear².

O precursor do Método Simplex foi Frank Lauren Hitchcock³ (1875-1957), em artigo de 1941 – como já comentamos anteriormente –, e que Dantzig com justiça reconhece.

Dantzig afirma que a inspiração para o método se origina do livro de Leontief “*The Structure of American Economy*”, publicado em 1932⁴. Dantzig formula o problema primal

$$\begin{array}{ll} \max & c_1 \cdot x_1 + \dots + c_n \cdot x_n \\ \text{sa} & a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n = b_1; \\ & \vdots \quad \quad \quad \vdots \\ & a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n = b_m; \\ & x_1 \geq 0, \dots, x_n \geq 0, \end{array}$$

e dedica o artigo para a construção de uma solução.

O nome “Método Simplex” resulta do fato de que eventualmente uma das equações no conjunto das restrições é $x_1 + \dots + x_n = 1$, que junto com $x_1 \geq 0, \dots, x_n \geq 0$ define o simplexo básico do $\mathbb{R}^n$. O Método Simplex é uma variação do usual método de Gauss de linha-redução.

Dantzig lembra que

[...] Vendo que os economistas não tinham um método de solução, decidi tentar minha própria sorte para encontrar um algoritmo. Tenho uma grande dívida para com Jerzy Neyman, o líder mundial da Estatística Matemática, que orientou meu trabalho de doutorado em Berkeley. Minha tese foi sobre dois famosos problemas não resolvidos em estatística matemática que eu resolvi interpretando equivocadamente como um dever para casa [...] Na terminologia de hoje, minha tese foi sobre a existência de multiplicadores de Lagrange (ou variáveis duais) para um programa linear sobre um continuum de variáveis cada uma delas limitada entre zero e um e satisfazendo restrições lineares expressas na forma de integrais de Lebesgue. Havia também um[a] [função] objetivo linear a ser otimizado[a]. A geometria particular usada na minha tese estava na dimensão das colunas em vez das linhas. Essa geometria das colunas deu o insight que me fez acreditar que o Método Simplex seria uma técnica de solução muito eficiente para resolver programas lineares, eu o propus no verão de 1947 e por sorte deu certo.⁵

O objetivo desta lição é estudar o Método Simplex de Dantzig.

2. O Método Simplex e a linha-redução de sistemas lineares

Se um sistema linear tem menos equações linearmente independentes que variáveis, o que chamamos “resolver” o sistema por linha-redução, caso seja possível, é expressar algumas variáveis, ditas a partir de então “dependentes”, em função das demais, chamadas “independentes”: “resolver” o sistema significa que podemos atribuir valores arbitrários às variáveis independentes e calcular, através das equações obtidas pela linha-redução, os valores das variáveis dependentes.

Exemplo 1: Seja o sistema

$$\begin{aligned}3 \cdot x_1 - 5 \cdot x_2 + 2 \cdot x_3 + 4 \cdot x_4 &= 1 \\6 \cdot x_1 - 10 \cdot x_2 - 4 \cdot x_3 - 6 \cdot x_4 &= 1.\end{aligned}$$

Por linha-redução, resulta

$$\left(\begin{array}{cccc|c}3 & -5 & 2 & 4 & 1 \\6 & -10 & -4 & -6 & 1\end{array}\right) \sim \left(\begin{array}{cccc|c}1 & -5/3 & 0 & 1/6 & 1/4 \\0 & 0 & 1 & 7/4 & 1/8\end{array}\right).$$

A solução é dada por

$$\begin{aligned}x_1 &= \frac{1}{4} + \frac{5}{3} \cdot x_2 - \frac{1}{6} \cdot x_4 \\x_3 &= \frac{1}{8} - \frac{7}{4} \cdot x_4.\end{aligned}$$

Observe que a linha-redução conduz automaticamente a expressar x_1 e x_3 como funções de x_2 e x_4 , e não x_1 e x_2 como funções de x_3 e x_4 ! É fácil ver porque isso ocorreu. Se pudéssemos expressar x_1 e x_2 como funções de x_3 e x_4 , então a matriz linha-reduzida teria a forma $\begin{pmatrix} 1 & 0 & b_{13} & b_{14} \\ 0 & 1 & b_{23} & b_{24} \end{pmatrix}$, o que implicaria que a submatriz $\begin{pmatrix} 3 & -5 \\ 6 & -10 \end{pmatrix}$ seria linha-equivalente à matriz $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ e, portanto, inversível! Entretanto, essa submatriz não é inversível, pois $\begin{vmatrix} 3 & -5 \\ 6 & -10 \end{vmatrix} = 0$. A linha-redução se dá sobre a submatriz $\begin{pmatrix} 3 & 2 \\ 6 & -4 \end{pmatrix}$, pois $\begin{vmatrix} 3 & 2 \\ 6 & -4 \end{vmatrix} = -24$.

Facilmente percebemos que as variáveis x_i e x_j podem ser expressas através das demais, se $\begin{vmatrix} a_{1i} & a_{1j} \\ a_{2i} & a_{2j} \end{vmatrix} \neq 0$. Por exemplo, $\begin{vmatrix} 2 & 4 \\ -4 & -6 \end{vmatrix} = 4$, então podemos expressar as variáveis x_3 e x_4 como funções das variáveis x_1 e x_2 :

Por linha-redução, resulta:

$$\left(\begin{array}{cccc|c}3 & -5 & 2 & 4 & 1 \\6 & -10 & -4 & -6 & 1\end{array}\right) \sim \left(\begin{array}{cccc|c}-21/2 & 35/2 & 1 & 0 & -5/2 \\6 & -10 & 0 & 1 & 3/2\end{array}\right).$$

A solução é dada por:

$$\begin{aligned}x_3 &= -\frac{5}{2} + \frac{21}{2} \cdot x_1 - \frac{35}{2} \cdot x_2 \\x_4 &= \frac{3}{2} - 6 \cdot x_1 + 10 \cdot x_2.\end{aligned}$$

Observe que, em princípio, podemos ter como variáveis dependentes $\{x_1, x_2\}$, $\{x_1, x_3\}$, $\{x_1, x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$ e $\{x_3, x_4\}$, ou seja, as $\binom{4}{2} = 6$ combinações possíveis de quatro variáveis tomadas duas a duas. A combinação

$\{x_1, x_2\}$ não é possível porque $\begin{vmatrix} 3 & -5 \\ 6 & -10 \end{vmatrix} = 0$. As demais combinações são possíveis, pois os demais determinantes $\begin{vmatrix} a_{1i} & a_{1j} \\ a_{2i} & a_{2j} \end{vmatrix} \neq 0$ são não-nulos.

Exercício 1

Seja o sistema:

$$\begin{aligned} 3 \cdot x_1 - 5 \cdot x_2 + 2 \cdot x_3 + 4 \cdot x_4 &= 1 \\ 6 \cdot x_1 - 10 \cdot x_2 + 4 \cdot x_3 - 6 \cdot x_4 &= 1. \end{aligned}$$

Observe que não podemos ter as combinações de variáveis $\{x_1, x_2\}$ e $\{x_1, x_3\}$ como variáveis dependentes. Por que? A linha-redução resulta em:

$$\left(\begin{array}{cccc|c} 3 & -5 & 2 & 4 & 1 \\ 6 & -10 & 4 & -6 & 1 \end{array} \right) \sim \left(\begin{array}{cccc|c} 1 & -5/3 & 2/3 & 0 & 5/21 \\ 0 & 0 & 0 & 1 & 1/14 \end{array} \right)$$

ou

$$\begin{aligned} x_1 &= \frac{5}{21} + \frac{5}{3} \cdot x_2 + \frac{2}{3} \cdot x_3 \\ x_4 &= \frac{1}{14}. \end{aligned}$$

Se temos um sistema linha-reduzido como:

$$\begin{aligned} x_1 - \frac{5}{3} \cdot x_2 + \frac{1}{6} \cdot x_4 &= \frac{1}{4} \\ x_3 + \frac{7}{4} \cdot x_4 &= \frac{1}{8}. \end{aligned}$$

Vamos expressar o sistema através de um dispositivo chamado “*tableau simplex*”:

As variáveis dependentes estão na primeira coluna, no caso $\{x_1, x_3\}$, chamemos de “básicas”; as restantes, ou seja, as variáveis independentes, estão na primeira linha, no caso $\{x_2, x_4\}$, que chamamos de “não-básicas”. Observe que se $x_2 = x_4 = 0$, então os valores das variáveis dependentes estão na última coluna, ou seja, $x_1 = 1/4$ e $x_3 = 1/8$.

Um sistema qualquer com m equações e n incógnitas que já esteja linha-reduzido nas variáveis básicas $x_{b_1}, x_{b_2}, \dots, x_{b_m}$

$$\begin{aligned}
x_{i1} + a_{h,j_1} \cdot x_{j_1} + a_{h,j_2} \cdot x_{j_2} + \dots + a_{h,j_{n-m}} \cdot x_{j_{n-m}} &= b_1 \\
x_{i2} + a_{h,j_1} \cdot x_{j_1} + a_{h,j_2} \cdot x_{j_2} + \dots + a_{h,j_{n-m}} \cdot x_{j_{n-m}} &= b_2 \\
&\vdots \\
x_{im} + a_{h,j_1} \cdot x_{j_1} + a_{h,j_2} \cdot x_{j_2} + \dots + a_{h,j_{n-m}} \cdot x_{j_{n-m}} &= b_m
\end{aligned}$$

é representado via o *tableau* como:

	x_{j_1}	x_{j_2}	$\dots$	$x_{j_{n-m}}$	
x_{i1}	a_{h,j_1}	a_{h,j_2}	$\dots$	$a_{h,j_{n-m}}$	b_1
x_{i2}	a_{h,j_1}	a_{h,j_2}	$\dots$	$a_{h,j_{n-m}}$	b_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{im}	a_{h,j_1}	a_{h,j_2}	$\dots$	$a_{h,j_{n-m}}$	b_m

Observe o *tableau* abaixo:

	x_3	x_4	
x_1	a	b	E
x_2	c	d	F

As variáveis básicas estão na primeira coluna, no caso $\{x_1, x_2\}$; as variáveis não-básicas estão na primeira linha, no caso $\{x_3, x_4\}$. Observe que se $x_3 = x_4 = 0$, então os valores das variáveis dependentes estão na última coluna, ou seja, $x_1 = E$ e $x_2 = F$. O *tableau* corresponde ao sistema:

$$\begin{aligned}
x_1 + a \cdot x_3 + b \cdot x_4 &= E \\
x_2 + c \cdot x_3 + d \cdot x_4 &= F.
\end{aligned}$$

Para tirarmos x_1 da base e colocarmos x_3 , por exemplo, agimos da seguinte maneira: dividimos a primeira equação por a , resultando $\frac{1}{a} \cdot x_1 + x_3 + \frac{b}{a} \cdot x_4 = \frac{E}{a}$; eliminamos x_3 na segunda equação, isto é, multiplicamos a equação acima por $-c$ e somamos na segunda equação resultando $-\frac{c}{a} \cdot x_1 + x_2 + \left(-\frac{c}{a} \cdot b + d\right) \cdot x_4 = -\frac{c}{a} \cdot E + F$. Agora, as variáveis básicas são x_2 e x_3 :

$$\begin{aligned}
\frac{1}{a} \cdot x_1 + x_3 + \frac{b}{a} \cdot x_4 &= \frac{E}{a}, \\
-\frac{c}{a} \cdot x_1 + x_2 + \left(-\frac{c}{a} \cdot b + d\right) \cdot x_4 &= -\frac{c}{a} \cdot E + F.
\end{aligned}$$

ou, ainda,

$$\begin{array}{cc}
x_1 & x_4
\end{array}$$

x_3	$\frac{1}{a}$	$\frac{b}{a}$	$\frac{E}{a}$
x_2	$-\frac{c}{a}$	$-\frac{c}{a} \cdot b + d$	$-\frac{c}{a} \cdot E + F$

Acompanhe a linha-redução abaixo e compare como a tabela acima:

$$\left(\begin{array}{cccc|c} 1 & 0 & a & b & E \\ 0 & 1 & c & d & F \end{array} \right) \sim \left(\begin{array}{cccc|c} \frac{1}{a} & 0 & 1 & \frac{b}{a} & \frac{E}{a} \\ 0 & 1 & c & d & F \end{array} \right) \sim \left(\begin{array}{cccc|c} \frac{1}{a} & 0 & 1 & \frac{b}{a} & \frac{E}{a} \\ -\frac{c}{a} & 1 & 0 & \left(d - \frac{c}{a} \cdot b\right) & \left(F - \frac{c}{a} \cdot E\right) \end{array} \right)$$

Vejamos como realizar esta operação diretamente sobre a tabela

	x_3	x_4	
x_1	a	b	E
x_2	c	d	F

Primeiramente, assinale as variáveis que serão trocadas, no caso x_1 por x_3 ; x_1 deixa de ser básica e x_3 passa a ser básica. O número a é chamado de pivô.

	x_3	x_4	
x_1	$\boxed{a}$	b	E
x_2	c	d	F

Em seguida, invertemos o pivô na posição em que se encontra, obtendo $\frac{1}{a}$, e no restante da linha do pivô multiplicamos por $\frac{1}{a}$, ou seja, dividimos por a . Assim, obtemos a primeira equação $\frac{1}{a} \cdot x_1 + x_3 + \frac{b}{a} \cdot x_4 = \frac{E}{a}$:

	x_1	x_4	
x_3	$\frac{1}{a}$	$\frac{b}{a}$	$\frac{E}{a}$

Na coluna em que está o pivô, multiplicamos por $\frac{1}{a}$, ou seja, dividimos por a , e trocamos o sinal para obter $-\frac{c}{a}$:

$$x_1$$

x_3	$\frac{1}{a}$
x_2	$-\frac{c}{a}$

Para obter o número na posição $\times$, multiplicamos $-\frac{c}{a}$ por b que está acima na linha do pivô na tabela inicial, e somamos em d na posição em que se encontra o $\times$:

	x_1	x_4	
x_3	$\frac{1}{a}$	$\frac{b}{a}$	$\frac{E}{a}$
x_2	$-\frac{c}{a}$	$\times$	

Resulta em:

	x_1	x_4	
x_3	$\frac{1}{a}$	$\frac{b}{a}$	$\frac{E}{a}$
x_2	$-\frac{c}{a}$	$-\frac{c}{a} \cdot b + d$	$\times \times$

Para obter o número na posição $\times \times$, procedemos da mesma maneira: multiplicamos E , que está acima na linha do pivô, por $-\frac{c}{a}$ e somamos em F :

	x_1	x_4	
x_3	$\frac{1}{a}$	$\frac{b}{a}$	$\frac{E}{a}$
x_2	$-\frac{c}{a}$	$-\frac{c}{a} \cdot b + d$	$-\frac{c}{a} \cdot E + F$

Exemplo 2: Voltemos ao sistema $\begin{matrix} 3 \cdot x_1 - 5 \cdot x_2 + 2 \cdot x_3 + 4 \cdot x_4 = 1 \\ 6 \cdot x_1 - 10 \cdot x_2 - 4 \cdot x_3 - 6 \cdot x_4 = 1. \end{matrix}$ Pela linha-redução, obtemos:

$$\begin{matrix} x_1 - \frac{5}{3} \cdot x_2 + \frac{1}{6} \cdot x_4 = \frac{1}{4} \\ x_3 + \frac{7}{4} \cdot x_4 = \frac{1}{8}. \end{matrix}$$

O *tableau* correspondente é:

	x_2	x_4	
x_1	$-\frac{5}{3}$	$\frac{1}{6}$	$\frac{1}{4}$
x_3	0	$\frac{7}{4}$	$\frac{1}{8}$

Vamos tirar a variável x_1 da base e no lugar por x_4 :

	x_2	x_4	
x_1	$-\frac{5}{3}$	$\boxed{\frac{1}{6}}$	$\frac{1}{4}$
x_3	0	$\frac{7}{4}$	$\frac{1}{8}$

Então:

	x_1	x_1	
x_4	$-\frac{5}{3} \times 6 = -10$	6	$\frac{1}{4}$
x_3	?	$-\frac{7}{4} \times 6 = -\frac{21}{2}$	?

	x_1	x_1	
x_4	-10	6	$\frac{1}{4}$
x_3	$\left(-\frac{21}{2}\right) \times \left(-\frac{5}{3}\right) + 0$	$-\frac{21}{2}$	$\left(-\frac{21}{2}\right) \times \left(\frac{1}{4}\right) + \frac{1}{8}$

ou

	x_1	x_1	
x_4	-10	6	$\frac{1}{4}$
x_3	$\frac{35}{2}$	$-\frac{21}{2}$	$-\frac{5}{2}$

Que é um resultado obtido anteriormente:

$$x_3 = -\frac{5}{2} + \frac{21}{2} \cdot x_1 - \frac{35}{2} \cdot x_2$$

$$x_4 = \frac{3}{2} - 6 \cdot x_1 + 10 \cdot x_2.$$

Exemplo 3: Dado o sistema:

$$\begin{aligned}2 \cdot x_1 - x_2 + 2 \cdot x_3 - x_4 + 3 \cdot x_5 &= 14 \\ x_1 + 2 \cdot x_2 + 3 \cdot x_3 + x_4 + 0 \cdot x_5 &= 5 \\ x_1 + 0 \cdot x_2 - 2 \cdot x_3 + 0 \cdot x_4 - 2 \cdot x_5 &= -10\end{aligned}$$

cuja a matriz ampliada é:

$$\left(\begin{array}{ccccc|c}2 & -1 & 2 & -1 & 3 & 14 \\ 1 & 2 & 3 & 1 & 0 & 5 \\ 1 & 0 & -2 & 0 & -2 & -10\end{array}\right),$$

por linha-redução, obtemos $\left(\begin{array}{ccccc|c}1 & 0 & 0 & -2/17 & -2/17 & -4/17 \\ 0 & 1 & 0 & 11/17 & -23/17 & -80/17 \\ 0 & 0 & 1 & -1/17 & 16/17 & 83/17\end{array}\right)$, ou seja,

$$\begin{aligned}x_1 - \frac{2}{17} \cdot x_4 - \frac{2}{17} \cdot x_5 &= -\frac{4}{17} \\ x_2 + \frac{11}{17} \cdot x_4 - \frac{23}{17} \cdot x_5 &= -\frac{80}{17} \\ x_3 - \frac{1}{17} \cdot x_4 + \frac{16}{17} \cdot x_5 &= \frac{83}{17}.\end{aligned}$$

Usando o *tableau*, o sistema linha-reduzido:

	x_4	x_5	
x_1	$-\frac{2}{17}$	$-\frac{2}{17}$	$-\frac{4}{17}$
x_2	$\frac{11}{17}$	$-\frac{23}{17}$	$-\frac{80}{17}$
x_3	$-\frac{1}{17}$	$\frac{16}{17}$	$\frac{83}{17}$

Troquemos as variáveis x_1 e x_4 entre si.

	x_4	x_5	
x_1	$\boxed{-\frac{2}{17}}$	$-\frac{2}{17}$	$-\frac{4}{17}$
x_2	$\frac{11}{17}$	$-\frac{23}{17}$	$-\frac{80}{17}$
x_3	$-\frac{1}{17}$	$\frac{16}{17}$	$\frac{83}{17}$

	x_1	x_5	
x_4	$-\frac{17}{2}$	$\left(-\frac{17}{2}\right) \times \left(-\frac{2}{17}\right) = 1$	$\left(-\frac{17}{2}\right) \times \left(-\frac{4}{17}\right) = 2$
x_2	$-\left(-\frac{17}{2}\right) \times \left(\frac{11}{17}\right) = \frac{11}{2}$		
x_3	$-\left(-\frac{17}{2}\right) \times \left(-\frac{1}{17}\right) = -\frac{1}{2}$		

	x_1	x_5	
x_4	$-\frac{17}{2}$	1	2
x_2	$\frac{11}{2}$	$\left(\frac{11}{2}\right) \times \left(-\frac{2}{17}\right) - \frac{23}{17} = -2$	$\left(\frac{11}{2}\right) \times \left(-\frac{4}{17}\right) - \frac{80}{17} = -6$
x_3	$-\frac{1}{2}$	$\left(-\frac{1}{2}\right) \times \left(-\frac{2}{17}\right) + \frac{16}{17} = 1$	$\left(-\frac{1}{2}\right) \times \left(-\frac{4}{17}\right) + \frac{83}{17} = 5$

	x_1	x_5	
x_4	$-\frac{17}{2}$	1	2
x_2	$\frac{11}{2}$	-2	-6
x_3	$-\frac{1}{2}$	1	5

O sistema está expresso como

$$\begin{aligned}
 -\frac{17}{2} \cdot x_1 + x_4 + x_5 &= 2 & x_4 &= 2 + \frac{17}{2} \cdot x_1 - x_5 \\
 \frac{11}{2} \cdot x_1 + x_2 - 2 \cdot x_5 &= -6 & \text{ou} & x_2 &= -6 - \frac{11}{2} \cdot x_1 + 2 \cdot x_5 \\
 -\frac{1}{2} \cdot x_1 + x_3 + x_5 &= 5 & x_3 &= 5 + \frac{1}{2} \cdot x_1 - x_5.
 \end{aligned}$$

Observe que se $x_1 = x_5 = 0$, então $x_4 = 2$, $x_2 = -6$ e $x_3 = 5$.

Exemplo 4: Consideremos a reta que está na interseção dos planos

$$\begin{aligned}
 2 \cdot x_1 + x_2 - 2 \cdot x_3 &= 14 \\
 x_1 - 2 \cdot x_2 - 3 \cdot x_3 &= -5.
 \end{aligned}$$

Por linha-redução,

$$\left(\begin{array}{ccc|c} 2 & 1 & -2 & 14 \\ 1 & -2 & -3 & -5 \end{array} \right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -\frac{7}{5} & \frac{23}{5} \\ 0 & 1 & \frac{4}{5} & \frac{24}{5} \end{array} \right).$$

A reta na interseção dos planos é:

$$\begin{aligned}
 x_1 &= \frac{23}{5} + \frac{7}{5} \cdot x_3 \\
 x_2 &= \frac{24}{5} - \frac{4}{5} \cdot x_3.
 \end{aligned}$$

Se fizermos $x_3=0$, obtemos o ponto $P(23/3, 24/5, 0)$ pertencente à reta em questão. Escrevendo a forma reduzida do sistema acima através do *tableau*, obtemos

$$\begin{array}{ccc} & x_3 & \\ x_1 & \boxed{\begin{array}{c} 7 \\ -5 \end{array}} & \frac{23}{3} \\ x_2 & \frac{4}{5} & \frac{24}{5} \end{array}$$

Por razão ilustrativa, digamos que retiramos x_1 da base, então obtemos

$$\begin{array}{ccc} & x_1 & \\ x_3 & -\frac{5}{7} & -\frac{115}{21} \\ x_2 & \frac{4}{7} & \frac{964}{105} \end{array}$$

Disso concluímos que o ponto $P(0, 964/105, -115/21)$ também está sobre a reta.

A primeira observação importante é que não queremos valores negativos para as variáveis básicas quando anulamos as variáveis não-básicas porque, em geral, nos problemas de Economia, os valores relevantes para as variáveis são não-negativos. Observe que no *tableau* inicial os valores das variáveis $x_1 = \frac{23}{3}$ e $x_2 = \frac{24}{5}$ são não-negativos quando fazemos $x_3 = 0$.

Portanto, se estivéssemos resolvendo um problema de Economia deveríamos ter descartado a substituição da variável x_1 por x_3 , pois essa troca devolve um valor negativo para x_3 quando fizermos $x_1 = 0$.

A primeira regra é que jamais escolhemos numa coluna um valor negativo ou nulo para determinar a linha do pivoteamento. Voltando ao *tableau* inicial, determinamos a linha correta para o *pivot*

$$x_3$$

x_1	$-\frac{7}{5}$	$\frac{23}{3}$
x_2	$\boxed{\frac{4}{5}}$	$\frac{24}{5}$

O resultado é:

	x_2	
x_1	$\frac{7}{4}$	$\frac{241}{15}$
x_3	$\frac{5}{4}$	6

O ponto $P(241/15, 0, 6)$ está sobre a reta e não tem coordenada negativa.

Exemplo 5: Consideremos a interseção dos dois hiperplanos

$$\begin{aligned}x_1 + 0 \cdot x_2 + 20 \cdot x_3 - 11 \cdot x_4 &= 3 \\ 0 \cdot x_1 + x_2 + 5 \cdot x_3 - 2 \cdot x_4 &= 1.\end{aligned}$$

Observe que o sistema está linha-reduzido nas variáveis x_1 e x_2 :

	x_3	x_4		θ
x_1	20	-11	3	$\theta = \frac{3}{20}$
x_2	5	-2	1	$\theta = \frac{1}{5}$

Por enquanto não dê importância para a última coluna. Se trocarmos alguma variável da base por outra que não esteja na base, então x_4 não seria escolhida. Por que? Resta x_3 e temos duas possibilidades: ou retiramos x_1 ou retiramos x_2 . Retirando x_1 da base:

x_3	x_4
x_1	$\boxed{20} \quad -11 \quad 3$
x_2	$5 \quad -2 \quad 1$

resulta

	x_1	x_4	
x_3	$\frac{1}{20}$	$-\frac{11}{20}$	$\frac{3}{20}$

$$x_2 \qquad \qquad \qquad -\frac{1}{4} \qquad \qquad \qquad -\frac{19}{4} \qquad \qquad \qquad \frac{1}{4}$$

Se $x_1 = x_4 = 0$, então $x_2 = \frac{1}{4}$ e $x_3 = \frac{3}{20}$ e os valores das variáveis são não-negativos. Se retiramos x_2 da base:

$$\begin{array}{ccc} & x_3 & x_4 \\ x_1 & 20 & -11 & 3 \\ x_2 & \boxed{5} & -2 & 1 \end{array}$$

resulta

$$\begin{array}{ccc} & x_3 & x_4 \\ x_1 & -4 & -33 & -1 \\ x_2 & \frac{1}{5} & -\frac{2}{5} & \frac{1}{5} \end{array}$$

Se $x_3 = x_4 = 0$, então $x_1 = -1$, $x_2 = \frac{1}{5}$ e como a variável x_1 exibe um valor negativo, pela razão acima, não estamos interessados em tais soluções. Como determinar a escolha entre x_1 e x_2 ?

Voltando ao *tableau*

$$\begin{array}{ccc} & x_3 & x_4 \\ x_1 & a & b & E \\ x_2 & c & d & F \\ & \uparrow & & \end{array}$$

Olhando a primeira coluna, em princípio, podemos substituir x_1 ou x_2 por x_3 . **Qual deve ser a escolha?** Se, por exemplo, substituirmos x_1 por x_3 , obtemos na última coluna $\frac{E}{a}$, que é o valor de x_3 se fizermos $x_1 = x_4 = 0$. **Se estamos procurando soluções não-negativas**, então, necessariamente, devemos ter $\frac{E}{a} \geq 0$. Digamos que $E \geq 0$, então, necessariamente, devemos ter $a > 0$. Portanto, escolhida a coluna do pivoteamento (no caso, a primeira), a linha escolhida está entre aquelas cujos elementos correspondentes são positivos. Aquelas linhas cujos elementos são zero ou negativos estão excluídas. Isso

havíamos concluído. Por essa razão, no exemplo, excluímos a coluna da variável x_4 .

Caso $a > 0$ e $c > 0$, podemos substituir x_1 ou x_2 por x_3 . **Qual deve ser a escolha?** Se estamos procurando **soluções não-negativas** para nossos sistemas, devemos escolher de tal modo que $-\frac{c}{a} \cdot E + F \geq 0$ no primeiro caso, isto é, x_1 por x_3 , ou que $E - \frac{a}{c} \cdot F \geq 0$ no segundo caso, isto é, x_2 por x_3 . Em outras palavras, $\frac{F}{c} \geq \frac{E}{a}$ ou $\frac{F}{c} \leq \frac{E}{a}$, isto é, escolhemos a troca na linha em que aparece a menor das razões entre $\frac{E}{a}$ e $\frac{F}{c}$. Chamemos $\theta = \frac{E}{a}$ e $\theta = \frac{F}{c}$ de razões teta. Escolhemos, portanto, a menor dessas razões.

Retornando ao exemplo. As razões são $\theta = \frac{3}{20}$ e $\theta = \frac{1}{5}$; se escolhermos a primeira linha onde está a menor razão θ , obtemos a última coluna com valores não-negativos; se escolhermos a segunda linha, onde está a maior razão θ , obtemos pelo menos um valor negativo.

Em resumo, determinada a coluna do *tableau*, a linha escolhida é aquela com a menor razão θ dentre as linhas com coeficientes positivos na coluna. Se existirem linhas com a mesma razão θ , e essa razão é a menor entre as razões, então é indiferente entre essas linhas qual escolhamos. Escolha qualquer uma entre elas!

3. Obtendo os vértices de polígonos

A primeira aplicação dos *tableaux* que mostramos é que eles são úteis para descobrir vértices de poliedros. Dantzig observa que

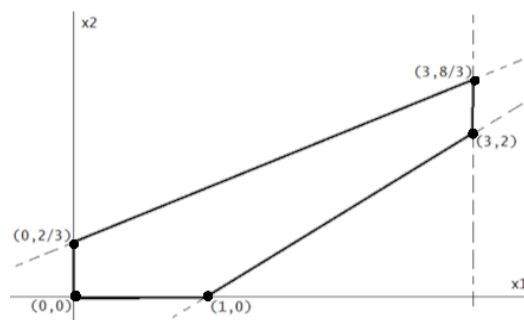
Durante o verão de 1947, Leonid Hurwicz, bem conhecido econometrista associado à Comissão Cowles, trabalhou com o autor em técnicas para resolver problemas de programação linear. Esse esforço e algumas sugestões de T. C. Koopmans resultaram no “Método Simplex”. Anteriormente, numa avaliação intuitiva, a ideia óbvia de mover-se de um vértice para o próximo ao longo das arestas de um poliedro convexo (que fundamenta o método simplex) tinha sido rejeitada como ineficiente. [Entretanto], numa geometria diferente, ela parecia eficiente e, por isso, felizmente, foi testada e aceita⁶.

O livro “*Linear Programming and Extensions*”, do qual tiramos o trecho acima, tem muitas informações históricas sobre o desenvolvimento da Programação Linear.

Exemplo 6: Consideremos o polígono definido pelas inequações abaixo:

$$\begin{aligned}x_1 - x_2 &\leq 1 \\ -2 \cdot x_1 + 3 \cdot x_2 &\leq 2 \\ x_1 + 0 \cdot x_2 &\leq 3 \\ x_1 \geq 0, x_2 &\geq 0.\end{aligned}$$

O polígono com os vértices está ilustrado abaixo:



Primeiramente, acrescentamos três variáveis $x_3 \geq 0, x_4 \geq 0$ e $x_5 \geq 0$, chamadas variáveis de folga, com o objetivo de obter equações a partir das inequações acima:

$$\begin{aligned}x_1 - x_2 + x_3 &= 1 \\ -2 \cdot x_1 + 3 \cdot x_2 + x_4 &= 2 \\ x_1 + 0 \cdot x_2 + x_5 &= 3 \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 &\geq 0.\end{aligned}$$

Observe que uma variável de folga, por exemplo x_3 em $x_1 - x_2 + x_3 = 1$, é a diferença $x_3 = 1 - (x_1 - x_2)$, justificando o nome “folga”.

O acréscimo das variáveis de folga produz um sistema linha-reduzido nessas variáveis:

	x_1	x_2		θ
x_3	1	-1	1	$\theta = 1$
x_4	-2	3	2	

x_5	1	0	3	$\theta = 3$
-------	---	---	---	--------------

Observe que se $x_1 = x_2 = 0$, ou seja, se estamos no vértice $(0,0)$ do polígono, então $x_3 = 1$, $x_4 = 2$ e $x_5 = 3$.

Podemos tirar algumas dessas variáveis da base, trocando-as por x_1 ou x_2 . Digamos que escolhamos x_1 . Não podemos trocar x_1 por x_4 , porque o coeficiente de x_1 é -2 na segunda equação; a escolha tem que ser entre x_3 e x_5 . Para não resultar variável com valor negativo, escolhemos a primeira linha, porque ali está a menor razão θ , ou seja, trocamos x_1 por x_3 . Isso significa que o valor de x_1 irá aumentar, ou seja, estaremos nos deslocando ao longo do eixo x_1 .

	x_1	x_2		θ
x_3	<u>1</u>	-1	1	$\theta = 1$
x_4	-2	3	2	
x_5	1	0	3	$\theta = 3$

Resulta

	x_3	x_2	
x_1	1	-1	1
x_4	2	1	4
x_5	-1	1	2

Se fizermos $x_2 = x_5 = 0$, então $x_1 = 1$, ou seja, estamos no vértice $(1,0)$.

Continuemos! Digamos que queremos pôr x_2 na base. Podemos trocar x_2 por x_1 , x_4 ou x_5 . A troca de x_2 por x_1 está eliminada, pois o coeficiente de x_2 é negativo na primeira equação. A menor razão θ está na terceira linha. Se trocarmos x_2 por x_5 estamos nos deslocando no sentido do eixo x_2 , ou seja, aumentando x_2 .

	x_3	x_2	θ
x_1	1	-1	1

x_4	2	1	4	$\theta = 4$
x_5	-1	<u>1</u>	2	$\theta = 2$

Resulta

	x_3	x_5	
x_1	0	1	3
x_4	3	-1	2
x_2	-1	1	2

Se fizermos $x_3 = x_5 = 0$, então $x_1 = 3$ e $x_2 = 2$, ou seja, estamos no vértice $(3, 2)$

Digamos que, no *tableau* acima, queremos pôr x_3 na base, então temos apenas uma alternativa, que é trocar x_3 por x_4 . Por que? O resultado é:

	x_4	x_5	
x_1	0	1	3
x_3	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$
x_2	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{8}{3}$

Fazendo $x_4 = x_5 = 0$, então $x_1 = 3$ e $x_2 = \frac{8}{3}$, ou seja, estamos no vértice $(3, 8/3)$.

Digamos que no *tableau* acima queremos retirar x_1 da base. Existe apenas uma possibilidade: trocar x_1 por x_5 . Não há problema, porque a menor razão θ na coluna do x_5 está na primeira linha:

	x_4	x_5	θ
x_1	0	<u>1</u>	3 $\theta = 3$
x_3	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$
x_2	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{8}{3}$ $\theta = 4$

Resulta,

x_4	x_1	θ
-------	-------	----------

x_5	0	1	3	$\theta = 3$
x_3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{5}{3}$	
x_2	$\frac{1}{3}$	$-\frac{2}{3}$	$\frac{2}{3}$	$\theta = 2$

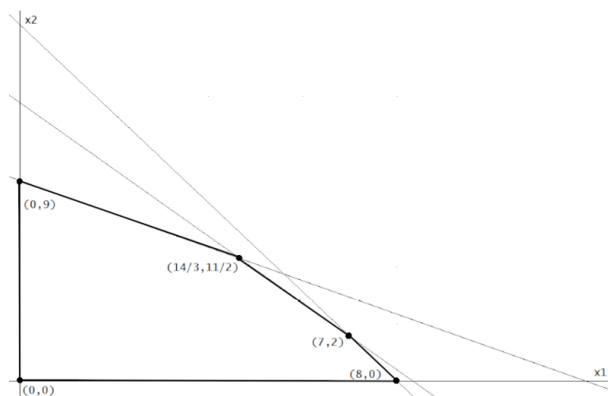
Fazendo $x_1 = x_5 = 0$, então $x_2 = \frac{2}{3}$, ou seja, estamos no vértice $(0, 2/3)$. Portanto, através dessas linha-reduções sucessivas, percorremos a partir do vértice $(0,0)$ todos os vértices do polígono! Este é o objetivo: obter os vértices do polígono!

Exercício 2

Usando os *tableaux*, encontre os vértices do polígono definido pelas inequações:

$$\begin{aligned} 3 \cdot x_1 + 4 \cdot x_2 &\leq 36 \\ 3 \cdot x_1 + 2 \cdot x_2 &\leq 25 \\ 2 \cdot x_1 + x_2 &\leq 16 \\ x_1 \geq 0, x_2 &\geq 0. \end{aligned}$$

Saia com a base x_3, x_4, x_5 e, sucessivamente, x_3, x_1, x_5 ; x_3, x_2, x_1 ; x_5, x_2, x_1 e x_5, x_2, x_4 .

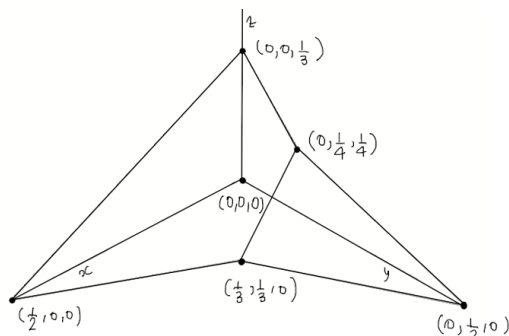


Exercício 3

Usando os *tableaux*, encontre os vértices do poliedro definido pelas inequações:

$$\begin{aligned} 2 \cdot x_1 + x_2 + 3 \cdot x_3 &\leq 1 \\ x_1 + 2 \cdot x_2 + 2 \cdot x_3 &\leq 1 \\ x_1 \geq 0, x_2 \geq 0, x_3 &\geq 0. \end{aligned}$$

Observando o *tableau* $\begin{array}{c|c|c|c|c} & x_1 & x_2 & x_3 & \\ \hline x_4 & 2 & 1 & 3 & 1 \\ \hline x_5 & 1 & 2 & 2 & 1 \end{array}$, note que não podem existir vértices neste poliedro, tal que $x_1 > 0, x_2 > 0, x_3 > 0$. Por quê?



Comentando sobre a teoria matemática dos jogos, Morgenstern observa que

[...] sua estrutura acaba sendo bem diferente daquela da mecânica clássica e, a fortiori, do cálculo diferencial e integral. Isso se deve ao caráter essencialmente combinatório dos problemas encontrados e à ampla divergência entre os fenômenos subjacentes e os fenômenos físicos.”⁷

Como já observamos anteriormente, von Neumann acreditava que a Teoria Econômica impunha a necessidade de uma Matemática adequada, que os métodos do cálculo são a Matemática criada para a Mecânica e que sua aplicação à Teoria Econômica apresentava inconvenientes. Um comentário que Morgenstern atribuiu a von Neumann, num artigo sobre o período em que escreviam “*Game Theory*”, foi muito mal interpretado à época, como se fosse uma crítica de Morgenstern a algum autor – talvez Hicks, provavelmente Samuelson – imputada a Neumann, que já estava morto. Entretanto, para além da suscetibilidade, a suposta frase de Neumann expressa o que de fato ele pensava, ou seja, que estavam construindo uma Matemática nova para a Economia:

Você sabe, Oskar, se esses livros forem desenterrados em algum momento daqui a algumas centenas de anos, as pessoas não acreditarão que foram escritos no nosso tempo. Ao contrário, as pessoas os acharão contemporâneos do Newton, tão primitiva é a

matemática. Simplesmente, a Economia está ainda a um milhão de milhas de distância do estado em que está uma ciência avançada tal como a física.⁸

Sublinhe-se “primitiva” para ser usada em Economia, pois a Economia aguardava a sua matemática!

4. Encontrando os pontos ótimos em programação linear

Exemplo 7: Acrescentemos uma função a ser otimizada. Seja o problema:

$$\begin{array}{ll} \max & 4 \cdot x_1 + 5 \cdot x_2 \\ & x_1 - x_2 \leq 1 \\ \text{sa} & -2 \cdot x_1 + 3 \cdot x_2 \leq 2 \\ & x_1 + 0 \cdot x_2 \leq 3 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Outra grande ideia atribuída a von Neumann para a resolução de um problema de programação linear é incorporar a função objetivo a ser otimizada como mais uma equação no sistema linear que resulta das restrições. Vejamos!

Seja $F = 4 \cdot x_1 + 5 \cdot x_2$, que é uma isoquanta da função objetivo. Vamos interpretar F como uma variável de folga e $-4 \cdot x_1 - 5 \cdot x_2 + F = 0$ como uma equação a ser acrescida no conjunto das restrições:

$$\begin{array}{l} x_1 - x_2 + x_3 = 1 \\ -2 \cdot x_1 + 3 \cdot x_2 + x_4 = 2 \\ x_1 + 0 \cdot x_2 + x_5 = 3 \\ -4 \cdot x_1 - 5 \cdot x_2 + F = 0 \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, \end{array}$$

cujo *tableau* é:

	x_1	x_2	
x_3	1	-1	1
x_4	-2	3	2
x_5	1	0	3
F	-4	-5	0

Uma condição que impomos é que jamais retiremos a variável F da base. Essa variável é a que queremos maximizar, portanto, é sempre dependente das demais! Jamais a trataremos como uma variável independente.

Observe que estamos no vértice $x_1 = x_2 = 0$, ou seja, estamos no vértice $(0,0)$ do polígono. Nesse vértice, $F = 0$. Como o objetivo é aumentar F , então vamos sair desse vértice. Avançamos na direção do eixo x_1 ou na direção do eixo x_2 ? O crescimento marginal de F na direção do eixo x_1 é 4 e na direção do eixo x_2 é 5, então avançamos na direção do eixo x_2 , pois nessa direção F cresce mais.

	x_1	x_2	
x_3	1	-1	1
x_4	-2	3	2
x_5	1	0	3
F	-4	-5	0
		↑	

Então,

	x_1	x_2	
x_3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{5}{3}$
x_4	$-\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
x_5	1	0	3
F	$-\frac{22}{3}$	$\frac{5}{3}$	$\frac{10}{3}$

Observe que avançamos para o vértice $(0, 2/3)$. Nesse vértice, $F = \frac{10}{3}$.

A última equação do sistema é $-\frac{22}{3} \cdot x_1 + \frac{5}{3} \cdot x_2 + F = \frac{10}{3}$, ou $F = \frac{10}{3} + \frac{22}{3} \cdot x_1 - \frac{5}{3} \cdot x_2$, portanto, se aumentarmos x_1 nos deslocando na direção do eixo x_1 , o valor de F aumenta, pois o coeficiente de x_1 é positivo. Logo, vamos pôr x_1 na base!

	x_1	x_4		θ
x_3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{5}{3}$	$\theta = 5$
x_2	$-\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	
x_5	$\boxed{1}$	0	3	$\theta = 2$
F	$-\frac{22}{3}$	$\frac{5}{3}$	$\frac{10}{3}$	
	$\uparrow$			

Resultando em:

	x_5	x_4	
x_3	$-\frac{1}{3}$		
x_2	$\frac{2}{3}$		$\frac{8}{3}$
x_1	1	0	3
F	$\frac{22}{3}$	$\frac{5}{3}$	$\frac{76}{3}$

A última equação do sistema é $\frac{22}{3} \cdot x_1 + \frac{5}{3} \cdot x_4 + F = \frac{76}{3}$, ou $F = \frac{76}{3} - \frac{22}{3} \cdot x_1 - \frac{5}{3} \cdot x_4$, portanto, se aumentarmos x_1 ou x_4 , estaremos diminuindo o valor de F . Sendo assim, quando na última linha existem apenas coeficientes positivos ou nulos, a função objetivo está maximizada e o valor máximo para F é valor na última coluna e na última linha, no caso, $F = \frac{76}{3}$. Enquanto existirem valores negativos nessa coluna, o valor de F pode ser aumentado.

Observe que não precisamos percorrer todo o polígono ou obter todos os vértices. Saímos da origem $(0,0)$ e avançamos percorrendo as arestas para vértices adjacentes ao vértice em que estamos não diminuindo o valor de F . Como o número de vértices é finito e sempre existe um caminho pelas arestas que conduz de qualquer vértice para qualquer outro vértice, necessariamente temos que alcançar o vértice onde está o maior valor de F . O método substitui a conferência dos valores da função objetivo em todos os vértices para

verificar onde estão os ótimos por um caminho que conduz a um ótimo.

Exemplo 8: Existe um resultado adicional muito importante.

O problema dual associado é

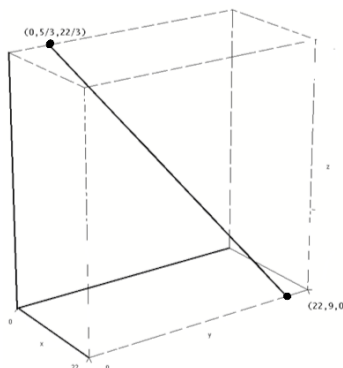
$$\begin{array}{ll} \min & y_1 + 2 \cdot y_2 + 3 \cdot y_3 \\ \text{sa} & y_1 - 2 \cdot y_2 + y_3 \geq 4 \\ & -y_1 + 3 \cdot y_2 + 0 \cdot y_3 \geq 5 \\ & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{array}$$

Os vértices da região estão sobre o segmento da reta na interseção dos dois planos que pertence ao primeiro octante $y_1 \geq 0, y_2 \geq 0, y_3 \geq 0$. A reta que está na interseção dos planos é:

$$\left(\begin{array}{ccc|c} 1 & -2 & 1 & 4 \\ -1 & 3 & 0 & 5 \end{array} \right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 3 & 22 \\ 0 & 1 & 1 & 9 \end{array} \right);$$

$$\begin{array}{l} y_1 = 22 - 3 \cdot t \\ y_2 = 9 - t \\ y_3 = 0 + t. \end{array}$$

Os extremos do segmento que são os vértices da região são $(22, 9, 0)$ e $(0, 5/3, 22/3)$. Veja o gráfico abaixo:



O mínimo da função objetivo $f(y_1, y_2, y_3) = y_1 + 2 \cdot y_2 + 3 \cdot y_3$ está em pelo menos um desses vértices. No caso, no vértice $(0, 5/3, 22/3)$.

Observe que o mínimo $f = \frac{76}{3}$ da função acima é o máximo da função $F(x_1, x_2) = 4 \cdot x_1 + 5 \cdot x_2$.

Se voltarmos ao último *tableau*:

	x_5	x_4	
x_3	$-\frac{1}{3}$		
x_2	$\frac{2}{3}$		$\frac{8}{3}$
x_1	1	0	3
F	$\frac{22}{3}$	$\frac{5}{3}$	$\frac{76}{3}$

Verificamos que os valores ótimos das variáveis de escolha, $y_1 \geq 0, y_2 \geq 0, y_3 \geq 0$, do problema dual, aparecem na última linha do *tableau*. Isso não é acaso!

5. A dualidade

Os valores ótimos $y_1^* = 0$, $y_2^* = \frac{5}{3}$ e $y_3^* = \frac{22}{3}$ são os coeficientes das variáveis de folga na última equação do *tableau*:

$$\underbrace{0}_{\uparrow y_1^*} \cdot x_3 + \underbrace{\frac{5}{3}}_{\uparrow y_2^*} \cdot x_4 + \underbrace{\frac{22}{3}}_{\uparrow y_3^*} \cdot x_5 + F = \underbrace{\frac{76}{3}}_{F^* = f^*}.$$

Esse resultado importante e intrigante é chamado na literatura de “Teorema da Dualidade” ou simplesmente “Dualidade”. O esclarecimento desse teorema terá que aguardar a continuação do curso.

Deve-se a **Von Neumann** a descoberta da dualidade. Este comunicou suas ideias para **Dantzig**, inclusive a ideia que um jogo poderia ser reduzido a um problema de programação linear, num encontro dos dois em Princeton em 1948:

Doze anos depois [1960], lembro-me de uma conversa com o professor Tucker que estava lendo o manuscrito de meu livro “*Linear Programming and Extensions*”. Nossa conversa foi assim: “Por que”, ele perguntou, “atribuíste a dualidade a von Neuman e não ao meu grupo?” “Porque ele foi o primeiro a mostrá-la para mim.” Tucker respondeu: “Isto é estranho, pois não encontramos nada escrito sobre o que von Neumann teria feito em relação à dualidade. O que temos é seu artigo “*On a Maximizing Problem*.” “Verdade”, eu disse, “mas deixe-me enviar-lhe um artigo que escrevi como

resultado de meu primeiro encontro com von Neumann”. Enviei-lhe meu relatório “*A Theorem on Linear Inequalities*”, datado de 5 de janeiro de 1948, que contém (até onde eu sei) a primeira prova rigorosa da dualidade. Mais tarde, Tucker me perguntou: “Por que você não publicou?” ao que respondi: “Porque o resultado não é meu – é de von Neumann.” Tudo o que fiz foi escrever, para circulação interna, minha própria prova daquilo que von Neumann esboçara. Esta era o meu jeito de educar as pessoas em meu escritório no Pentágono. Hoje, todos citam Von Neumann como o criador do teorema da dualidade e creditam a Tucker, Kuhn e Gale a publicação da primeira prova rigorosa.⁹

Exemplo 9: Consideremos o problema da dieta:

$$\begin{array}{ll} \min & 2 \cdot y_1 + y_2 \\ \text{sa} & 3 \cdot y_1 + y_2 \geq 3 \\ & 4 \cdot y_1 + 3 \cdot y_2 \geq 6 \\ & y_1 + 3 \cdot y_2 \geq 2 \\ & y_1 \geq 0, y_2 \geq 0. \end{array}$$

Resolvemos este problema através problema primal associado:

$$\begin{array}{ll} \max & 3 \cdot x_1 + 6 \cdot x_2 + 2 \cdot x_3 \\ \text{sa} & 3 \cdot x_1 + 4 \cdot x_2 + x_3 \leq 2 \\ & x_1 + 3 \cdot x_2 + 3 \cdot x_3 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{array}$$

	x_1	x_2	x_3		θ
x_4	3	4	1	2	$\theta = \frac{1}{2}$
x_5	1	3	3	1	$\theta = \frac{1}{3}$
F	-3	-6	-2	0	
		↑			

	x_1	x_5	x_3		θ
x_4	$\frac{5}{3}$	$-\frac{4}{3}$	-3	$\frac{2}{3}$	$\theta = \frac{2}{5}$
x_2	$\frac{1}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\theta = 1$
F	-1	2	4	2	
	↑				

	x_4	x_5	x_3
x_1	$\frac{3}{5}$		

x_2	$-\frac{1}{5}$			
F	$\frac{3}{5}$	$\frac{6}{5}$	$\frac{11}{5}$	$\frac{12}{5}$

Portanto, $y_1^* = \frac{3}{5}$, $y_2^* = \frac{6}{5}$ e o valor mínimo de $f(y_1, y_2) = 2 \cdot y_1 + y_2$ é $f^* = \frac{12}{5}$.

O Teorema da Dualidade foi publicado por David Gale (1921-2008), Albert Tucker (1905-1995) e Harold Kuhn (1925-2014) no encontro número zero da programação linear¹⁰.

Exercício 4

Usando o **Método Simplex** resolva o problema:

$$\begin{array}{ll} \max & 2 \cdot x + y + z \\ \text{sa} & x + y + 2 \cdot z \leq 18 \\ & -2 \cdot x + y - z \geq -6 \\ & x \geq 0, y \geq 0, z \geq 0. \end{array}$$

Exercício 5

Usando a dualidade e o Método Simplex, resolva o problema de minimização (problema da dieta):

$$\begin{array}{ll} \min & 20 \cdot x_1 + 60 \cdot x_2 \\ \text{sa} & 2 \cdot x_1 + 4 \cdot x_2 \geq 3 \\ & 5 \cdot x_1 + 2 \cdot x_2 \geq 1 \\ & 5 \cdot x_1 + x_2 \geq 5 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Exercício 6

Os alimentos A, B e C possuem os nutrientes I e II. As informações sobre A, B e C estão na tabela abaixo:

Nutrientes	Alimento A	Alimento B	Alimento C	Quantidades diárias necessárias
I	4 unidades	9 unidades	7 unidades	36 unidades
II	1 unidades	1 unidades	3 unidades	10 unidades

Em relação ao alimento A, o gasto diário de uma unidade é R\$ 6, em relação ao alimento B, o gasto diário de uma unidade é R\$ 3 e, em relação ao alimento C, o gasto diário de uma unidade é R\$ 4.

Usando a dualidade, resolva o problema acima indicando como devemos combinar os alimentos A, B e C de maneira a satisfazer as necessidades nutricionais ao menor custo. Explícite o problema que está sendo resolvido.

Exercício 7

Uma empresa química dispõe de 30 ton de nitrato e 80 ton de fosfato para produzir três tipos diferentes de fertilizantes. Na tabela abaixo estão os dados de uso e receita para cada um dos fertilizantes.

tipo de fertilizante	nitrato	fosfato	Receita
I	2 ton/1000 sacos	4 ton/1000 sacos	R\$ 30 /1000 sacos
II	3 ton/1000 sacos	4 ton/1000 sacos	R\$ 60 /1000 sacos
III	2 ton/1000 sacos	2 ton/1000 sacos	R\$ 50 /1000 sacos

Quanto de cada tipo de fertilizante a empresa deve produzir se o objetivo for maximizar a receita? Use o **Método Simplex**.

Exercício 8

Considere o jogo

$$\begin{pmatrix} -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Formule os problemas do jogador das linhas e do jogador das colunas. Resolva pelo Método Simplex um desses problemas e utilize a dualidade para resolver o outro. Qual é o equilíbrio de Nash desse jogo? A solução pode ser obtida facilmente por inspeção da matriz.

Exercício 9

Considere o jogo chamado “*rock-scissors-paper*”

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Formule os problemas do jogador das linhas e do jogador das colunas. Resolva pelo **Método Simplex** um desses problemas e utilize a **dualidade** para resolver o outro. Qual é o **equilíbrio de Nash** desse jogo?

6. Usando o Método Simplex para resolver o problema dual diretamente

A essência do Método Simplex está em percorrer os vértices adjacentes de um poliedro através da aresta que esses determinam. A simplicidade desse princípio é responsável pela flexibilidade do método, que permite adequações para muitas situações. Não existe nada que impossibilite que o empreguemos para resolver o problema dual diretamente.

Exemplo 10: Seja o problema:

$$\begin{array}{ll} \min & 6 \cdot x_1 + 4 \cdot x_2 \\ \text{sa} & 4 \cdot x_1 + x_2 \geq 8 \\ & 9 \cdot x_1 + x_2 \geq 20 \\ & 7 \cdot x_1 + 3 \cdot x_2 \geq 14 \\ & 10 \cdot x_1 + 40 \cdot x_2 \geq 40 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

que transformamos no problema:

$$\begin{array}{ll} \max & -6 \cdot x_1 - 4 \cdot x_2 \\ \text{sa} & -4 \cdot x_1 - x_2 + x_3 = -8 \\ & -9 \cdot x_1 - x_2 + x_4 = -20 \\ & -7 \cdot x_1 - 3 \cdot x_2 + x_5 = -14 \\ & -10 \cdot x_1 - 40 \cdot x_2 + x_6 = -40 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0 \end{array}$$

Ou seja, transformamos num problema de maximização e acrescentamos as variáveis de folga x_3, x_4, x_5, x_6 . O *tableau* deste problema é:

	x_1	x_2		
x_3	-4	-1	-8	$\theta = 8$
x_4	-9	<u>-1</u>	-20	$\theta = 20 \leftarrow$
x_5	-7	-3	-14	$\theta = \frac{14}{3}$
x_6	-10	-40	-40	$\theta = 1$
F	6	$\frac{4}{\dagger}$	0	

«A primeira coisa a observar é que $F = -6 \cdot x_1 - 4 \cdot x_2$ na margem aumenta mais na direção do eixo x_2 , então nesse problema começamos pela coluna dois. Como as variáveis $x_i \geq 0$, mas os coeficientes são negativos, procuramos pela linha onde está a **maior razão** θ para que não resultem números negativos na última coluna. O *tableau* resultante é:

	x_1	x_4		
x_3	5	-1	12	$\theta = \frac{12}{5}$
x_2	9	-1	20	$\theta = \frac{20}{9}$
x_5	20	-3	46	$\theta = \frac{46}{20}$
x_6	<u>350</u>	-0	760	$\theta = \frac{76}{35} \leftarrow$
F	$-\frac{30}{\dagger}$	4	0	-80

A partir do segundo *tableau* passamos a agir como anteriormente, procurando aumentar na direção que oferece maior acréscimo marginal e na linha que exhibe o menor θ :

	x_6	x_4	
x_3			
x_2	$-\frac{9}{350}$		$\frac{16}{35}$
x_5			
x_1	$\frac{1}{350}$		$\frac{76}{35}$
F	$\frac{3}{35}$	$\frac{4}{7}$	$-\frac{104}{7}$

Portanto, a solução ótima para o problema acima é $x_1^* = \frac{76}{35}$ e $x_2^* = \frac{16}{35}$ e o valor mínimo é $\frac{104}{7}$. A solução do problema primal

$$\begin{aligned}
 \max \quad & 8 \cdot y_1 + 20 \cdot y_2 + 14 \cdot y_3 + 40 \cdot y_4 \\
 \text{sa} \quad & 4 \cdot y_1 + 9 \cdot y_2 + 7 \cdot y_3 + 10 \cdot y_4 \leq 6 \\
 & y_1 + y_2 + 3 \cdot y_3 + 40 \cdot y_4 \leq 4 \\
 & y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, y_4 \geq 0,
 \end{aligned}$$

está resolvida pela dualidade: os valores ótimos são os coeficientes das variáveis de folga: $y_1^* = 0, y_2^* = \frac{4}{7}, y_3^* = 0, y_4^* = \frac{3}{35}$, como nos problemas anteriores.

Exercício 10

Considere o jogo $\begin{pmatrix} 2 & 2 & -1 \\ 2 & -1 & 1 \end{pmatrix}$. Formule os problemas do jogador das linhas e do jogador das colunas. Não é difícil ver que a solução para o problema das linhas é $x_1^* = \frac{1}{3}, x_2^* = \frac{1}{2}$ e para o problema das colunas é $y_1^* = 0, y_2^* = \frac{1}{3}, y_3^* = \frac{1}{2}$. Resolva pelo Método Simplex o problema das linhas e, consequentemente, o problema das colunas usando o *tableau* abaixo:

	x_1	x_2	
x_3	-3	-3	-1
x_4	-3	0	-1
x_5	0	-2	-1
F	$\frac{1}{3}$	1	0

e utilize a dualidade para resolver o outro. Qual é o equilíbrio de Nash desse jogo? Qual é o valor desse jogo?

Exercício 11

Certa empresa tem capacidade para produzir três produtos. Para produzir uma unidade por hora do produto I são necessárias 5 unidades do insumo 1 e 10 unidades do insumo 2. Para produzir uma unidade por hora do produto II são necessárias 3 unidades do insumo 1 e 8 unidades do insumo 2. Para produzir uma unidade por hora do produto III são necessárias 5 unidades do insumo 1 e 15 unidades do insumo 2. Estão disponíveis, por hora, 15 unidades do insumo 1 e 32 unidades do insumo 2. Uma unidade de I é vendida por R\$ 2, uma unidade de II é vendida por R\$ 4 e uma unidade de III é vendida por R\$ 1. Encontre os preços sombra que correspondem aos insumos resolvendo diretamente o problema dual.

	y_1	y_2			y_4	y_1	
y_3	-5	-10	-2	y_3			
y_4	-3	-8	-4	y_2	$-\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$
y_3	-5	-15	-1	y_3			
F	15	32		F	4	3	-16

Quais são as quantidades ótimas dos três produtos? Qual é a receita?

Exercício 12

Determinada empresa necessita de uma liga metálica especial que contenha no mínimo 3% de chumbo, 3% de zinco e 2% de estanho. No mercado, existem à venda duas ligas, digamos I e II, medidas numa unidade particular para esse tipo de material. As especificações dessas ligas estão na tabela abaixo

Liga	chumbo	zinco	estanho	Preço em reais por unidade
I	2%	3%	5%	3
II	5%	4%	1%	2

A empresa resolve adquirir as ligas existentes no mercado e fundi-las, de modo a obter a mistura que necessita. Usando a dualidade, resolva o problema acima indicando como devemos combinar as ligas I e II de maneira que uma unidade da liga requerida seja obtida ao menor custo. Explícite o problema que está sendo resolvido. (Esse problema é uma versão simplificada de outro problema formulado por Dantzig¹¹.)

Exemplo 11: Existem problemas no Método Simplex? Existem! Obviamente, nosso objetivo não é conduzir um estudo exaustivo de Programação Linear. Isso é impossível aqui. A Programação Linear é uma área da Matemática que obviamente exige um estudo mais prolongado e profundo para conhecê-la de maneira aceitável.

7. O exemplo de Beale

O exemplo a seguir foi elaborado pelo matemático inglês Evelyn M. L. Beale¹² (1928-1985), com o objetivo de mostrar que existem situações em que o Método Simplex fica “preso” em “loops” sem obter o ponto ótimo. Seguimos a apresentação de Louis Brickman¹³ no livro “*Mathematical Introduction to Linear Programming and Game Theory*”, do qual adotamos a forma de representar os *tableaux*.

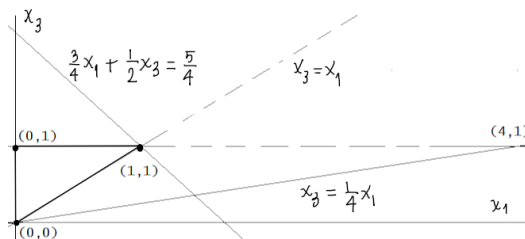
Seja o problema:

$$\begin{aligned} \max \quad & \frac{3}{4} \cdot x_1 - 20 \cdot x_2 + \frac{1}{2} \cdot x_3 - 6 \cdot x_4 \\ \text{sa} \quad & \frac{1}{4} \cdot x_1 - 8 \cdot x_2 - x_3 + 9 \cdot x_4 \leq 0 \\ & \frac{1}{2} x_1 - 12 \cdot x_2 - \frac{1}{2} \cdot x_3 + 3 \cdot x_4 \leq 0 \\ & 0 \cdot x_1 + 0 \cdot x_2 + x_3 + 0 \cdot x_4 \leq 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0. \end{aligned}$$

É fácil perceber que o valor ótimo para as variáveis x_2 e x_3 é zero. Com isso, o problema torna-se:

$$\begin{aligned} \max \quad & \frac{3}{4} \cdot x_1 + \frac{1}{2} \cdot x_3 \\ \text{sa} \quad & \frac{1}{4} \cdot x_1 - x_3 \leq 0 \\ & \frac{1}{2} x_1 - \frac{1}{2} \cdot x_3 \leq 0 \\ & 0 \cdot x_1 + x_3 \leq 1 \\ & x_1 \geq 0, x_3 \geq 0. \end{aligned}$$

A região determinada pela restrição está desenhada abaixo e, observando-a, concluímos que a solução ótima está no vértice $x_1 = x_3 = 1$. Portanto, a solução do problema inicial é $x_1^* = 1, x_2^* = 0, x_3^* = 1, x_4^* = 0$. O valor máximo é $\frac{5}{4}$.



Abaixo estão os *tableaux* para a resolução do problema:

	x_1	x_2	x_3	x_4			x_3	x_2	x_3	x_4	
x_5	$\boxed{\frac{1}{4}}$	-8	-1	9	0	x_1	4	-32	-4	36	0
x_6	$\frac{1}{2}$	-12	$-\frac{1}{2}$	3	0	x_6	-2	$\boxed{4}$	$\frac{3}{2}$	-15	0
x_7	0	0	1	0	1	x_7	0	0	1	0	1
F	$-\frac{3}{4}$	20	$-\frac{1}{2}$	6	0	F	3	-4	$-\frac{7}{2}$	33	0

	x_5	x_6	x_3	x_4			x_5	x_6	x_1	x_4	
x_1	-12	8	$\boxed{8}$	-84	0	x_3	$-\frac{3}{2}$	1	$\frac{1}{8}$	$-\frac{21}{2}$	0
$\rightarrow x_2$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{8}$	$-\frac{15}{4}$	0	x_2	$\frac{1}{16}$	$-\frac{1}{8}$	$-\frac{3}{64}$	$\frac{3}{16}$	0
x_7	0	0	1	0	1	x_7	$\frac{3}{2}$	-1	$-\frac{1}{8}$	$\frac{21}{2}$	1
F	1	1	-2	18	0	F	-2	3	$\frac{1}{4}$	-3	0

	x_5	x_6	x_1	x_2			x_3	x_6	x_1	x_2	
x_3	$\boxed{2}$	-6	$-\frac{5}{2}$	56	0	x_5	$\frac{1}{2}$	-3	$-\frac{5}{4}$	28	0
$\rightarrow x_4$	$\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{4}$	$\frac{16}{3}$	0	x_4	$-\frac{1}{6}$	$\boxed{\frac{1}{3}}$	$\frac{1}{6}$	-4	0
x_7	-2	6	$\frac{5}{2}$	-56	1	x_7	1	0	0	0	1
F	-1	1	$-\frac{1}{2}$	16	0	F	$\frac{1}{2}$	-2	$-\frac{7}{4}$	44	0

	x_3	x_4	x_1	x_2	
x_5	-1	9	$\frac{1}{4}$	-8	0
$\rightarrow x_6$	$-\frac{1}{2}$	3	$\frac{1}{2}$	-12	0
x_7	1	0	0	0	1
F	$-\frac{1}{2}$	6	$-\frac{3}{4}$	20	0

Esse último *tableau* é o primeiro *tableau*! Voltamos ao início. Estamos num *loop* que não conduz ao ótimo. Mas não se preocupe! Isso não acontece em nossos exercícios.

8. Dantzig e o Método Simplex

A maneira como expomos o Método Simplex não é a mesma que Dantzig o apresenta no artigo original. A maneira acima se origina de um artigo de 1954, do matemático americano Carlton Eduard

Lemke¹⁴ (1920-2004). Esquematizamos de maneira breve a demonstração de como Dantzig¹⁵ formulou o problema primal:

$$\begin{array}{ll} \max & c_1 \cdot x_1 + \dots + c_n \cdot x_n \\ \text{sa} & a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n = b_1; \\ & \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ & a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n = b_m; \\ & x_1 \geq 0, \dots, x_n \geq 0. \end{array}$$

Na forma matricial, representaremos o problema acima como:

$$\begin{array}{ll} \max & c' \cdot x \\ \text{sa} & A \cdot x = b; \\ & x \geq 0; \end{array}$$

e que $c' = (c_1 \ \dots \ c_n)$, $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$ e $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \geq 0$.

Quando acrescentamos as variáveis de folga ao sistema de desigualdades que definem o poliedro, obtemos como restrição um sistema linear $A \cdot x = b$ como acima. Notemos por $P_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{R}^m$, $j = 1, \dots, n$, cada coluna da matriz A e $P_0 = b$. Na notação de Dantzig, o problema é encontrar $\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \geq 0$, tal que $\lambda_1 \cdot P_1 + \dots + \lambda_n \cdot P_n = P_0$, tal que $\lambda_1 \cdot c_1 + \dots + \lambda_n \cdot c_n = \max z$, ou seja, $A \cdot \lambda = b$, $\lambda \geq 0$, $\lambda_1 \cdot c_1 + \dots + \lambda_n \cdot c_n = \max z$. Um vetor $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \geq 0$, tal que $A \cdot x = b$, é dito um vetor viável e se, além de viável, também satisfaz $x_1 \cdot c_1 + \dots + x_n \cdot c_n = \max z$, é dito uma solução viável maximizadora.

Dantzig parte de dois pressupostos. O primeiro é que todas as $\binom{n}{m}$ combinações possíveis $\{P_{j_1}, \dots, P_{j_m}\}$ de colunas de A são linearmente independentes no $\mathbb{R}^m$. Essa é uma suposição que objetiva tornar a demonstração do método mais simples e pode ser abandonada depois. Portanto, por exemplo, $\{P_1, \dots, P_m\}$ é uma base no $\mathbb{R}^m$. A segunda suposição é que exista um vetor viável $\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} > 0$, isto é, $\lambda_j > 0, j = 1, \dots, m$ e $\lambda_1 \cdot P_1 + \dots + \lambda_m \cdot P_m = P_0$. Essa suposição implica que existam pontos no interior do

poliedro e todos os problemas que tratamos satisfazem essa suposição. Notemos $\lambda_1 \cdot c_1 + \dots + \lambda_m \cdot c_m = z_0$.

Partindo dessas suposições, Dantzig demonstrou a existência de uma solução ótima, caso o conjunto dos vetores viáveis $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ determine um conjunto limitado de valores $c_1 \cdot x_1 + \dots + c_n \cdot x_n$.

Como $\{P_1, \dots, P_m\}$ é uma base no $\mathbb{R}^m$, então existem $x_{ij} \in \mathbb{R}$, não necessariamente $x_{ij} \geq 0$, tal que $x_{1j} \cdot P_1 + x_{2j} \cdot P_2 + \dots + x_{mj} \cdot P_m = P_j, j=1, \dots, n$. Digamos que $x_{1j} \cdot c_1 + x_{2j} \cdot c_2 + \dots + x_{mj} \cdot c_m = z_j, j=1, \dots, n$. Seja $\theta > 0$, então $(\lambda_1 - \theta \cdot x_{1j}) \cdot P_1 + (\lambda_2 - \theta \cdot x_{2j}) \cdot P_2 + \dots + (\lambda_m - \theta \cdot x_{mj}) \cdot P_m + \theta \cdot P_j = P_0$ e $(\lambda_1 - \theta \cdot x_{1j}) \cdot c_1 + (\lambda_2 - \theta \cdot x_{2j}) \cdot c_2 + \dots + (\lambda_m - \theta \cdot x_{mj}) \cdot c_m + \theta \cdot z_j = z_0$. Somando $\theta \cdot c_j$ na última equação resulta $(\lambda_1 - \theta \cdot x_{1j}) \cdot c_1 + (\lambda_2 - \theta \cdot x_{2j}) \cdot c_2 + \dots + (\lambda_m - \theta \cdot x_{mj}) \cdot c_m + \theta \cdot c_j = z_0 + \theta \cdot (c_j - z_j)$. Como cada $\lambda_j > 0, j=1, \dots, m$, então existe um intervalo $(0, \theta_0) \subseteq \mathbb{R}$, tal que $\theta \in (0, \theta_0)$, então $(\lambda_k - \theta \cdot x_{kj}) \geq 0$, para $k=1, \dots, m$, independentemente do sinal de $x_{ij}, i=1, \dots, m$.

Suponhamos ainda que exista $c_j > z_j$, então $z_0 + \theta_0 \cdot (c_j - z_j) > z_0$. Portanto, nossa conclusão é que se existe $c_j > z_j$, então $z_0 + \theta_0 \cdot (c_j - z_j) > z_0$, ou seja, podemos aumentar o valor z_0 através de vetores viáveis. Observe que se $x_{ij} \leq 0$, então $\lambda_k - \theta \cdot x_{kj} \geq \lambda_k > 0$, para todo $\theta > 0$. Portanto, a limitação de θ está determinada por aqueles coeficientes x_{ij} que são positivos.

Suponhamos que entre os $x_{ij}, i=1, \dots, m$ exista pelo menos um $x_{ij} > 0$. Como $\lambda_k - \theta \cdot x_{kj} \geq 0$, então $\frac{\lambda_k}{x_{kj}} \geq \theta$, se $x_{kj} > 0$. Se escolhermos $0 < \theta_0 = \min_{i_0 > 0} \frac{\lambda_k}{x_{kj}} = \frac{\lambda_{k_0}}{x_{k_0 j}}$, então $(\lambda_k - \theta_0 \cdot x_{kj}) \geq 0$. Nos casos em que $x_{kj} \leq 0$, então $\lambda_k - \theta \cdot x_{kj} \geq \lambda_k > 0$ para todo $\theta > 0$ e em particular para θ_0 . Desse modo, nossa conclusão é que se temos um vetor $\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} > 0$, tal que $\lambda_j > 0, j=1, \dots, m$ e $\lambda_1 \cdot P_1 + \dots + \lambda_m \cdot P_m = P_0$, então é possível aumentar $\lambda_1 \cdot c_1 + \dots + \lambda_m \cdot c_m = z_0$ se existe $c_j > z_j$.

Se todos os $x_{ij} \leq 0, i=1, \dots, m$, então θ não está limitado e podemos fazer $\theta \rightarrow +\infty$ e $z_0 + \theta \cdot (c_j - z_j) \rightarrow +\infty$, e neste caso teríamos um vetor viável com $m+1$ componentes e tal que $z_0 + \theta \cdot (c_j - z_j) \rightarrow +\infty$. Portanto, nesse caso, o problema é tal que o máximo não existe. Sendo assim, se o problema é limitado, que é o caso em todos os nossos exemplos, o processo acima não pode nos conduzir para algum $c_j > z_j$ e $x_{ij} \leq 0, i=1, \dots, m$.

Suponhamos, portanto, que o problema seja limitado, então se $c_j > z_j$, necessariamente existe $x_{kj} > 0$. Nesse caso, obtivemos um vetor viável $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \geq 0$, cujas coordenadas não-negativas são determinadas por $(\lambda_1 - \theta_0 \cdot x_{1j}) \cdot P_1 + (\lambda_2 - \theta_0 \cdot x_{2j}) \cdot P_2 + \dots + (\lambda_m - \theta_0 \cdot x_{mj}) \cdot P_m + \theta_0 \cdot P_j = P_0$, tal que $(\lambda_1 - \theta_0 \cdot x_{1j}) \cdot c_1 + (\lambda_2 - \theta_0 \cdot x_{2j}) \cdot c_2 + \dots + (\lambda_m - \theta_0 \cdot x_{mj}) \cdot c_m + \theta_0 \cdot c_j > z_0$. Observe que o número de componentes positivas de x é igual ou menor que m . Se eliminarmos P_{k_0} , porque $x_{k_0} = \lambda_{k_0} - \theta_0 \cdot x_{k_0j} = 0$, e o substituímos por P_j , pois $\theta_0 > 0$, então estamos com uma nova base do $\mathbb{R}^m$ $z_0 + \theta_0 \cdot (c_j - z_j) > z_0$. Se os coeficientes restantes $\lambda_i - \theta_0 \cdot x_{ij}$, $i \neq k_0$ forem positivos, então os tomamos junto com $\theta_0 > 0$ como um novo conjunto de λ 's positivos e reiniciamos o processo com a base formada pelos $P_1, \dots, P_m$, exceto P_{k_0} , que foi substituído por P_j . Enquanto isso for possível, estaremos em cada etapa aumentando $c_1 \cdot x_1 + \dots + c_n \cdot x_n = z$ a partir do valor inicial z_0 .

Num número finito de etapas, o processo se interrompe, pois são $\binom{n}{m}$ combinações possíveis $\{P_1, \dots, P_m\}$ de colunas de A .

Caso o conjunto das restrições seja limitado, temos duas possibilidades. Uma possibilidade é que $c_j \leq z_j$ para $j=1, \dots, n$. Nesse caso, então a etapa final nos fornece uma solução ótima. Vejamos!

Digamos que $\mu_1 \cdot P_1 + \dots + \mu_n \cdot P_n = P_0$, $\mu_j \geq 0$ e que $\mu_1 \cdot c_1 + \dots + \mu_n \cdot c_n = z$, ou seja, z é um valor qualquer possível para $c_1 \cdot x_1 + \dots + c_n \cdot x_n$. Por hipótese, $c_j \leq z_j$, então, $z = \mu_1 \cdot c_1 + \dots + \mu_n \cdot c_n \leq \mu_1 \cdot z_1 + \dots + \mu_n \cdot z_n$. Como $x_{1j} \cdot P_1 + x_{2j} \cdot P_2 + \dots + x_{mj} \cdot P_m = P_j, j=1, \dots, n$, então $\left(\sum_{j=1}^n \mu_j \cdot x_{1j} \right) \cdot P_1 + \dots + \left(\sum_{j=1}^n \mu_j \cdot x_{mj} \right) \cdot P_m = P_0$ e $z \leq \left(\sum_{j=1}^n \mu_j \cdot x_{1j} \right) \cdot c_1 + \dots + \left(\sum_{j=1}^n \mu_j \cdot x_{mj} \right) \cdot c_m$. Como $\lambda_1 \cdot P_1 + \dots + \lambda_m \cdot P_m = P_0$, pela independência linear dos $P_1, \dots, P_m$, então $\sum_{j=1}^n \mu_j \cdot x_{ij} = \lambda_i$. Portanto, $\lambda_1 \cdot c_1 + \dots + \lambda_m \cdot c_m \geq z$. Se numa das etapas $c_j \leq z_j$, para $j=1, \dots, n$, então atingimos o máximo do problema nessa etapa.

A outra maneira de o processo ser interrompido com algum $c_j > z_j$ é que existam coeficientes nulos $\lambda_k - \theta_0 \cdot x_{kj}$ em $(\lambda_1 - \theta_0 \cdot x_{1j}) \cdot P_1 + (\lambda_2 - \theta_0 \cdot x_{2j}) \cdot P_2 + \dots + (\lambda_m - \theta_0 \cdot x_{mj}) \cdot P_m + \theta_0 \cdot P_j = P_0$, além de $x_{k_0} = \lambda_{k_0} - \theta_0 \cdot x_{k_0j} = 0$. Nesse caso, eliminamos P_{k_0} e colocamos P_j , então diminuimos um pouco $0 < \theta_1 < \theta_0$, tal que $(\lambda_1 - \theta_1 \cdot x_{1j}) \cdot P_1 + (\lambda_2 - \theta_1 \cdot x_{2j}) \cdot P_2 + \dots + (\lambda_m - \theta_1 \cdot x_{mj}) \cdot P_m + \theta_1 \cdot P_j = P_0$ e cada coeficiente $\lambda_k - \theta_1 \cdot x_{kj} > 0$. Temos

uma nova base $\{P_1, \dots, P_m, P_j\}$, com m vetores, da qual se retirou P_k . Nesse caso, podemos reiniciar o processo a partir dessa base.

9. George Bernard Dantzig (1914-2005)

O pai de George Dantzig, Tobias Dantzig, também foi matemático. Tobias escreveu um livro de divulgação, "*Number, The Language of Science*" (1930), muito popular e que em 2007 ainda estava sendo publicado.

George Dantzig concluiu o bacharelado em matemática em 1936 e o mestrado em 1938. Depois de concluir o mestrado, Dantzig foi trabalhar no *Bureau of Labor Statistics*, de onde saiu para fazer doutorado em Berkeley sob supervisão do famoso estatístico Jerzy Neyman (1894-1981), onde destacou-se como matemático. Em 1941, a guerra interrompeu o doutorado de Dantzig, adiando a conclusão para 1946. Durante e depois da guerra, Dantzig trabalhou para a Força Aérea Norte Americana, o que foi decisivo para a sua carreira, pois foram os problemas de planejamento da Força Aérea que o conduziram à Programação Linear e ao Método Simplex. Em 1952, Dantzig passou a trabalhar na RAND Corporation e, em 1960, retornou para Berkeley como professor no Departamento de Engenharia Industrial. Em 1966, foi para Stanford onde permaneceu até sua aposentadoria em 1985.

Quando, em 1975, Koopmans e Kantorovich receberam o Prêmio Nobel de Economia, a comunidade científica ficou um tanto decepcionada, pois parecia injusto excluir Dantzig, afinal os trabalhos de Kantorovich e Dantzig apresentam similaridades. Conta-se que Koopmans hesitou em receber o prêmio porque achava injusta a exclusão de Dantzig¹⁶. Mas Dantzig recebeu inúmeros prêmios. No mesmo ano de 1975, recebeu o Prêmio von Neumann

para Teoria e a Medalha Nacional de Ciências, importantes honrarias científicas norte americanas.¹⁷

10. A Pesquisa Operacional (OR)

Durante a 2ª Guerra, a reunião de engenheiros, físicos, matemáticos, estatísticos, economistas e administradores em torno de objetivos comuns que exigiam um trabalho multidisciplinar, resultou numa nova disciplina chamada Pesquisa Operacional. A Pesquisa Operacional, para Philip M. Morse (1903-1985), é uma modalidade de engenharia, pois está preocupada com as aplicações para fins de resolução de problemas práticos e que exijam a combinação dos conhecimentos específicos das diferentes especializações acima. Se as diversas modalidades de Engenharia se preocupam com máquinas e a Economia e a Administração se preocupam com a organização racional dos recursos físicos e humanos, a Pesquisa Operacional se preocupa com a articulação de homens e máquinas. Nas palavras de Morse:

a Pesquisa Operacional não se preocupa com as máquinas ou com homens individualmente, mas com a operação como um todo; com táticas de batalha, com planejamento estratégico e logístico para operações futuras, com a inter-relação entre flutuações de vendas, tamanho dos estoques e programação da produção, com o padrão de fluxo de mercadoria em um grupo de fábricas ou tráfego em uma cidade, para citar alguns exemplos. Usando a palavra “nível” para distinguir entre os diferentes assuntos, se pudermos separar a palavra de qualquer conotação de importância ou dificuldade relativa. A física e a química corresponderiam então à pesquisa em nível básico ou material, a pesquisa em nível de engenharia ou aplicada corresponderia ao estudo de pontes e televisores. A pesquisa operacional seria então pesquisa no nível operacional.¹⁸

Inicialmente, a Pesquisa Operacional surgiu para resolver problemas relacionados à guerra. O índice do relatório “*Methods of Operations Research*”, elaborado por Morse e pelo químico George Kimball (1901-1967) para o Centro de Análise Naval, em 1946, não

deixa dúvida. Por exemplo, as seções dos capítulos tratam da busca aérea por submarinos, de combates aéreos, de combates entre comboios marítimos e submarinos etc.

Na versão americana, a Pesquisa Operacional baseou-se em Programação Linear, Teoria dos Jogos, Estatística e Simulações de Monte Carlo. A Teoria dos Jogos, por exemplo, está implicada na questão das patrulhas aéreas em busca de submarinos. O avião busca localizar o submarino e o último, evadir-se. É um jogo de soma-zero em que os oponentes escolhem estratégias mistas. A Estatística e simulações de Monte Carlo são usadas porque experimentos são necessários antes que grandes mudanças sejam implementadas.

Depois da guerra, Morse foi o principal proponente do uso civil da Pesquisa Operacional. Philip Morse foi físico e professor no MIT; de 1940 a 1945 foi diretor de pesquisa em sonares para a *National Defense Research Council and US Navy* e diretor do grupo de avaliação civil do *US Antisubmarine Program*. O relatório “*Methods of Operations Research*”¹⁹, elaborado por Morse e Kimball, depois que o sigilo foi retirado, tornou-se o primeiro livro sobre OR e publicado em 1951.

A Programação Linear e a Teoria dos Jogos que estamos estudando já deixaram bem claras suas aplicações militares e a dependência dessas em relação ao suporte civil. O artigo de Marshall Wood e George Dantzig não deixa dúvida:

No planejamento de programas militares, é necessário introduzir quantitativamente as várias limitações de recursos que restringem as capacidades militares em tempos de guerra, bem como em tempos de paz. Na maior parte, isso pode ser atribuído a limitações da economia industrial da nação. É preciso saber que parte da produção nacional total pode ser disponibilizada para fins militares. Isso não pode ser medido apenas em termos da capacidade produtiva da indústria aeronáutica ou da indústria de munições, assim como a força de uma Força Aérea não pode ser medida apenas em termos do número de grupos [de combate]. É necessário conhecer em detalhes as capacidades das indústrias de aço, alumínio, energia elétrica, transporte, mineração,

química e uma infinidade de outras indústrias de apoio às indústrias aeronáutica, naval e de munições, assim como é necessário conhecer as capacidades das atividades de treinamento, transporte, manutenção e abastecimento de apoio aos grupos aéreos de combate. Além disso, é necessário determinar se essas indústrias (ou atividades de apoio) estão balanceadas nas proporções adequadas para atender às mudanças de requisitos. Assim, uma vez que a determinação do “melhor” programa necessariamente começa com a consideração das limitações de recursos, ela deve necessariamente começar com a consideração das inter-relações das indústrias na economia industrial da nação.²⁰

Essas são evidências que levaram o historiador e economista Philip Morowski a tentar determinar os vestígios – se é que existem – que a coordenação dos interesses militares e civis durante a guerra deixaram na Teoria Econômica norte-americana do pós-guerra. Para Mirowski, a Teoria Econômica neoclássica adotou o conceito de racionalidade oriunda da lógica computacional e passou a usá-la para modelar o comportamento humano e de organizações. Aquilo que Mirowski chama de “*Cyborg Sciences*”²¹. A estatística e a econometria passaram a ser instrumentos de gerenciamento para as grandes empresas e para a administração pública. Além disso, a relação de assessoria que os técnicos civis e militares engendraram, em que os primeiros informam, mas não comandam e não têm responsabilidades sobre o curso dos acontecimentos, tornou-se a forma como os técnicos passaram a se relacionar com os dirigentes que os empregam.

Mirowski amplia o domínio de influência daquilo que outros autores já estavam enfatizando:

A geração de físicos que amadureceu com o nascimento da mecânica quântica possuía traços que permitiram não só se adaptar às condições existentes, mas, de fato, dominar e transformar o ambiente científico e tecnológico. [...] Estes desenvolvimentos após a Segunda Guerra Mundial foram profundamente influenciados pelas atividades de pesquisa e desenvolvimento realizadas durante as hostilidades, atividades em que os físicos desempenharam um papel dominante. Os padrões de interação que foram desenvolvidos pelos grupos OR entre cientistas e militares nos Estados Unidos tornaram-se o modelo para os comitês do pós-guerra que assessoraram as Forças Armadas no

desenvolvimento de armas e em questões táticas e estratégicas. Nos EUA, a Segunda Guerra Mundial foi responsável por um novo quadro de assessoria especializada através de cientistas civis para as forças armadas - principalmente físicos ou ex-físicos – como indicava a proliferação de comitês consultivos. [...] Enquanto os físicos e matemáticos aplicados foram predominantemente responsáveis pela criação da OR, economistas, advogados e cientistas sociais, assumiram as posições influentes na análise de sistemas. A análise eficaz dos sistemas deve integrar uma grande variedade de habilidades profissionais, incluindo as dos cientistas sociais e advogados. Os economistas e os advogados podem atuar como “generalistas e integradores” porque, por um lado, os aspectos mais técnicos dos sub-problemas podem ser relegados aos engenheiros e cientistas aplicados e, por outro lado, os aspectos mais integradores do problema - tais como a alocação de recursos entre os sistemas de armas competitivas caras, avaliar o custo de diferentes modos de pesquisa e desenvolvimento, negociar entre competidores, avaliar as conseqüências políticas de certas recomendações – foram as áreas tradicionais em que os kits de ferramentas e as habilidades especiais de economistas e advogados poderiam ser efetivamente empregadas.²²

Morowski cita um trecho da autobiografia de Morse:

Nenhum de nós foi o mesmo em relação ao que fomos em 1939 e suspeitávamos que nossas vidas não seria um retorno aos dias anteriores à guerra. Gostemos ou não, a física e os físicos transformaram a guerra e teriam inevitavelmente uma profunda influência nos tempos de paz.²³

¹ KOOPMANS, T. (ed.). **Activity Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951.

² DANTZIG, G. “Reminiscences About the Origins of Linear Programming”. **Operations Research Letters**, v. 1, nº 2 (1982), p. 45.

³ HITCHCOCK, F. L. “The Distribution of a Product from Several Sources to Numerous Localities”. **Journal of Mathematics and Physics**, nº 20 (1941), pp. 224-230.

⁴ DANTZIG, G. “Reminiscences About the Origins of Linear Programming”. **Operations Research Letters**, v. 1, nº 2 (1982), p. 45.

⁵ DANTZIG, G. “Reminiscences About the Origins of Linear Programming”. **Operations Research Letters**, v. 1, nº 2 (1982), pp. 45-46.

⁶ DANTZIG, G. **Linear Programming and Extensions**. California: The Rand Corporation, 1963, p. 24.

⁷ MORGENSTERN, O. “Game Theory”. In: WIENER, P. (ed.). **Dictionary of the History of Ideas, Studies of Select Pivotal Ideas**, v. II. New York: Charles Scribner’s Sons, 1973, p. 266.

- ⁸ MORGENSTERN, O. "The Collaboration Between Oskar Morgenstern and John von Neumann on the Theory of Game". *Journal of Economic Literature*, v. 14, n° 3 (1976), p. 810.
- ⁹ DANTZIG, G. "Reminiscences About the Origins of Linear Programming". **Operations Research Letters**, v. 1, n° 2 (1982), pp. 45-46.
- ¹⁰ GALE, D.; KUHN, H.; TUCKER, A. "Linear Programming and Theory of Games". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, pp. 317-329.
- ¹¹ DANTZIG, G. **Linear Programming and Extensions**. California: Rand Corporation, 1963, p. 43.
- ¹² BEALE, E. "Cycling in the Dual Simplex Algorithm". **Naval Research Logistic Quarterly**, v. 2, n° 4 (1955), pp. 269-275.
- ¹³ BRICKMAN, L. **Mathematical Introduction to Linear Programming and Game Theory**. New York: Springer Verlag, 1989.
- ¹⁴ LEMKE, C. E. "The Dual Method of Solving the Linear Programming Problem". **Naval Research Logistic Quarterly**, v. 1, n° 1 (1954), pp. 36-47.
- ¹⁵ DANTZIG, G. "Maximization of a Linear Function of Variables Subject to Linear Inequalities". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, pp. 339-347.
- ¹⁶ DÜPPE, T.; WEINTRAUB, E. "Siting the New Economic Science: The Cowles Commission's Activity Analysis Conference of June 1949". **Science in Context**, v. 27, n° 3 (2014), p. 456.
- ¹⁷ COTTLE, R. "George Dantzig: a Legendary Life in Mathematical Programming". **Mathematical Programming**, n° 105 (2006), pp. 1-8.
- ¹⁸ MORSE, P. "Operations Research". **Communications on Pure and Applied Mathematics**, v. VIII, 1955, p. 2.
- ¹⁹ MORSE, P.; KIMBALL, G. **Methods of Operations Research**. New York: Dover Publications, INC., 2003.
- ²⁰ DANTZIG, G.; WOOD, M. "The Programming of Interdependent Activities: A General Discussion". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, p. 18.
- ²¹ MIROWSKI, P. **Machine Dreams, Economics Becomes a Cyborg Science**. New York: Cambridge University Press, 2002.
- ²² SCHWEBER, S.; FORTUN, M. "Scientists and the Legacy of World War II: The Case of Operations Research (OR)". **Social Studies of Science**, v. 23, n° 4 (1993), p. 596.
- ²³ MIROWSKI, P. "Cyborg Agonistes: Economics Meets Operations Research in Mid-Century". **Social Studies of Science**, v. 29, n° 5 (1999), p. 69.

Lição QUATRO

Superfícies e Lagrangeanos

Sumário

Lição 4 – Superfícies e Lagrangeanos

1. Introdução
2. Superfícies de dimensão em
3. Superfícies de dimensão em
4. Superfícies tangentes
5. Lagrangeanos

Lição 4

Superfícies e Lagrangeanos

1. Introdução

Nosso objetivo nesta lição é tratar o problema de minimização ou maximização de uma função objetivo f definida sobre um domínio D de variáveis $x = (x_1, x_2, \dots, x_n)$, no qual as variáveis $x_1, x_2, \dots, x_n$ não são independentes umas em relação às outras e entre as quais a dependência é determinada por um conjunto de restrições $g_1(x_1, x_2, \dots, x_n) = c_1, \dots, g_m(x_1, x_2, \dots, x_n) = c_m$.

O domínio das variáveis determinado pelas restrições $g_1(x_1, x_2, \dots, x_n) = c_1, \dots, g_m(x_1, x_2, \dots, x_n) = c_m$ designamos, sob certas condições, como uma “superfície” de dimensão de $n-m$. Em cada ponto $P(x_1^*, x_2^*, \dots, x_n^*)$ sobre a superfície, definimos um plano, chamado plano tangente à superfície nesse ponto, que é fundamental para a teoria da otimização. Igualmente importante é o plano ortogonal ao plano tangente. As isoquantas da função $f=c$ também definem superfícies.

O resultado fundamental é que num ponto de mínimo ou máximo local de uma função f existe uma isoquanta da função f tangenciando a “superfície” determinada pela restrição. Esse resultado corresponde às condições de primeira ordem do lagrangeano e foi formulado por Lagrange no século 18.

Nesta lição é importante que se compreenda que a geometria acima, implicada na otimização, evidencie a razão pela qual as

condições de primeira ordem do lagrangeano são condições necessárias, mas não suficientes, para um mínimo ou máximo em problemas de otimização restrita. Em outras palavras, existir uma isoquanta da f tangenciando a restrição em determinado ponto não assegura que esse ponto seja um mínimo ou máximo de f sujeita à restrição

Ainda que as condições de primeira ordem não sejam, em geral, condições suficientes, nas próximas lições não abandonaremos o lagrangeano, mas, ao contrário, focamos nossa análise sobre as condições que, impostas às superfícies, implicam que as condições de primeira ordem do lagrangeano sejam suficientes para mínimos e máximos.

2. Superfícies de dimensão $(n-1)$ em $\mathbb{R}^n$

Seja uma isoquanta $f(x_1, x_2, \dots, x_n) = c$ da função f . Dizemos que $f(x_1, x_2, \dots, x_n) = c$ é uma superfície de dimensão $(n-1)$ no $\mathbb{R}^n$ se em cada ponto $(x_1^*, x_2^*, \dots, x_n^*)$, tal que $f(x_1^*, x_2^*, \dots, x_n^*) = c$, possamos definir um plano que chamamos de plano tangente à superfície no ponto $(x_1^*, x_2^*, \dots, x_n^*)$. Em outras palavras, estamos reservando a palavra “superfície” para os conjuntos de pontos do $\mathbb{R}^n$, definidos por equações $f(x_1, x_2, \dots, x_n) = c$, tais que em cada ponto $(x_1, x_2, \dots, x_n)$, $f(x_1, x_2, \dots, x_n) = c$ possamos definir um plano tangente. A seguir, veremos como se define o plano tangente.

Uma curva no $\mathbb{R}^n$ é uma função $\vec{x}: (t_0, t_1) \rightarrow \mathbb{R}^n$, $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_n(t)) \in \mathbb{R}^n$, tal que em cada ponto $t \in (t_0, t_1)$ existe $\vec{x}'(t) = \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{pmatrix}$. O vetor $\vec{x}'(t)$ é chamado de vetor tangente à curva, ou vetor velocidade, em t . Em geral, escreveremos esse vetor de maneira transposta, como $\vec{x}'(t) = (x_1'(t), x_2'(t), \dots, x_n'(t))$.

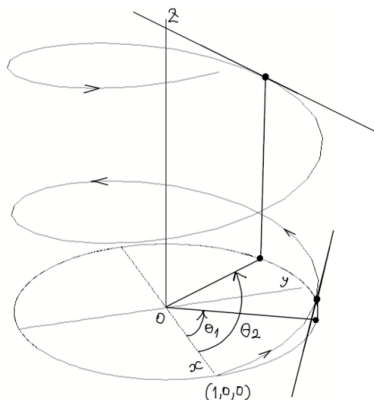
Neste estudo, nosso interesse em curvas se dá porque elas nos permitem definir vetores e retas tangentes às superfícies.

Exemplo 1: Seja o cilindro $f(x, y, z) = x^2 + y^2 = 1$. Uma curva sobre esse cilindro é a espiral $c(t) = (\cos(2 \cdot \pi \cdot t), \sin(2 \cdot \pi \cdot t), t)$, pois $f(\cos(2 \cdot \pi \cdot t), \sin(2 \cdot \pi \cdot t), t) = (\cos(2 \cdot \pi \cdot t))^2 + (\sin(2 \cdot \pi \cdot t))^2 = 1$. Derivando essa curva, obtemos $\vec{c}'(t) = (-2 \cdot \pi \cdot \sin(2 \cdot \pi \cdot t), 2 \cdot \pi \cdot \cos(2 \cdot \pi \cdot t), 1)$, que é um vetor tangente à curva em cada ponto $P(\cos(2 \cdot \pi \cdot t), \sin(2 \cdot \pi \cdot t), t)$. A reta tangente, por exemplo, no ponto t^* , parametrizada pela variável s , é dada por

$$c(t^*) + s \cdot \vec{c}'(t^*) = (\cos(2 \cdot \pi \cdot t^*), \sin(2 \cdot \pi \cdot t^*), t^*) + s \cdot (-2 \cdot \pi \cdot \sin(2 \cdot \pi \cdot t^*), 2 \cdot \pi \cdot \cos(2 \cdot \pi \cdot t^*), 1)$$

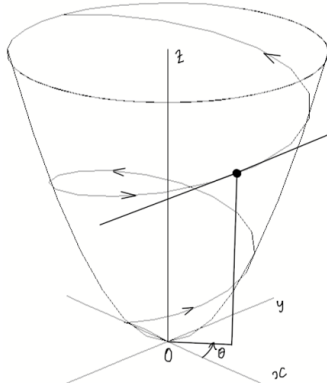
ou seja, a reta

$$\begin{aligned} x(s) &= \cos(2 \cdot \pi \cdot t^*) - 2 \cdot \pi \cdot s \sin(2 \cdot \pi \cdot t^*) \\ y(s) &= \sin(2 \cdot \pi \cdot t^*) + 2 \cdot \pi \cdot s \cos(2 \cdot \pi \cdot t^*) \\ z(s) &= t^* + s. \end{aligned}$$



No gráfico acima desenhamos essas retas tangentes para $t^* = \frac{1}{6}$ e $\frac{4}{3}$, que correspondem aos ângulos $\theta_1 = \frac{\pi}{3}$ e $\theta_2 = \frac{8 \cdot \pi}{3}$.

Exemplo 2: Consideremos a função $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, tal que $f(x, y) = x^2 + y^2$, cujo gráfico é o usual paraboloide abaixo. O gráfico dessa função é uma superfície, pois $z = x^2 + y^2$, que é a isoquanta $x^2 + y^2 - z = 0$ de nível 0 da função $F(x, y, z) = x^2 + y^2 - z = 0$. Abaixo temos um gráfico da superfície $x^2 + y^2 - z = 0$:

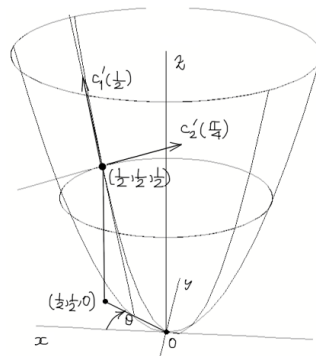


Nessa superfície definimos a curva $x(t) = (\sqrt{t} \cdot \cos(4 \cdot \pi \cdot t), \sqrt{t} \cdot \sin(4 \cdot \pi \cdot t), t)$, que está desenhada acima. Os vetores tangentes à curva são:

$$x'(t) = \left(\frac{\cos(4 \cdot \pi \cdot t)}{2 \cdot \sqrt{t}} - 4 \cdot \pi \cdot \sqrt{t} \cdot \sin(4 \cdot \pi \cdot t), \frac{\sin(4 \cdot \pi \cdot t)}{2 \cdot \sqrt{t}} + 4 \cdot \pi \cdot \sqrt{t} \cdot \cos(4 \cdot \pi \cdot t), 1 \right).$$

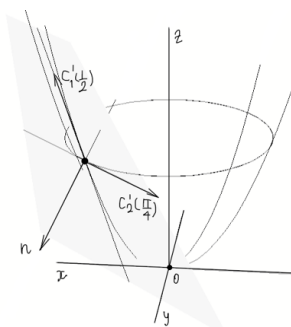
A reta tangente num ponto t_0 e parametrizada por s é dada por $r(s) = x(t_0) + s \cdot x'(t_0)$.

Na mesma superfície $x^2 + y^2 - z = 0$, considere as curvas $c_1(t) = (t, t, 2 \cdot t^2)$ e $c_2(s) = \left(\frac{\cos s}{\sqrt{2}}, \frac{\sin s}{\sqrt{2}}, \frac{1}{2} \right)$; os vetores tangentes a essas curvas são $c'_1(t) = (1, 1, 4 \cdot t)$ e $c'_2(s) = \left(\frac{-\sin s}{\sqrt{2}}, \frac{\cos s}{\sqrt{2}}, 0 \right)$, respectivamente. No ponto $P\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, $c'_1\left(\frac{1}{2}\right) = (1, 1, 2)$ e $c'_2\left(\frac{\pi}{4}\right) = \left(-\frac{1}{2}, \frac{1}{2}, 0\right)$; esses vetores estão desenhados abaixo:



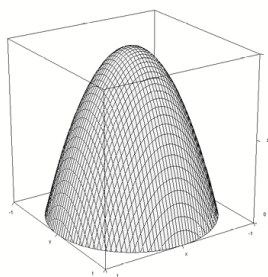
O plano tangente à superfície no ponto $P\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ é o plano que contém esses dois vetores e passa pelo ponto $P\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. Não é muito difícil ver que o vetor $n = (1, 1, -1)$ é perpendicular ao $c'_1\left(\frac{1}{2}\right) = (1, 1, 2)$ e ao $c'_2\left(\frac{\pi}{4}\right) = \left(-\frac{1}{2}, \frac{1}{2}, 0\right)$ e, portanto, é perpendicular ao plano que tangencia a superfície no

ponto $P\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, sendo assim, esse plano é $1 \cdot \left(x - \frac{1}{2}\right) + 1 \cdot \left(y - \frac{1}{2}\right) + (-1) \cdot \left(z - \frac{1}{2}\right) = 0$, ou seja, $x + y - z = \frac{1}{2}$. Veja o desenho abaixo:

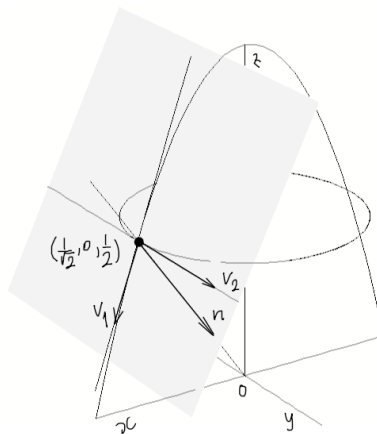


Encontrar o plano tangente não é muito prático. Em breve veremos que a determinação desse plano é extremamente simples.

Exemplo 3: Consideremos a função $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, tal que $f(x, y) = 1 - x^2 - y^2$, cujo gráfico é o usual parabolóide abaixo. O gráfico dessa função é uma superfície, pois $z = 1 - x^2 - y^2$ é a isoquanta $x^2 + y^2 + z = 1$ de nível 1 da função $F(x, y, z) = x^2 + y^2 + z = 1$.



Consideremos as curvas $c_1(t) = (t, 0, 1 - t^2)$ e $c_2(s) = \left(\frac{\cos s}{\sqrt{2}}, \frac{\sin s}{\sqrt{2}}, \frac{1}{2}\right)$ que estão sobre a superfície e se interceptam no ponto $P\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{2}\right)$. Os vetores tangentes às curvas acima são $c_1'(t) = (1, 0, -2t)$ e $c_2'(s) = \left(-\frac{\sin s}{\sqrt{2}}, \frac{\cos s}{\sqrt{2}}, 0\right)$, respectivamente. No ponto $P\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{2}\right)$, esses vetores são $c_1'\left(\frac{1}{\sqrt{2}}\right) = (1, 0, -\sqrt{2}) = v_1$ e $c_2'(0) = \left(0, -\frac{1}{\sqrt{2}}, 0\right) = v_2$. Esses vetores tangenciam as curvas e também o parabolóide. Um vetor ortogonal a ambos é $n = (-\sqrt{2}, 0, -1)$, portanto, o plano tangente à superfície no ponto $P\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{2}\right)$ é $\sqrt{2} \cdot \left(x - \frac{1}{\sqrt{2}}\right) - (z - 1/2) = 0$, ou $\sqrt{2} \cdot x - z = \frac{1}{2}$.



Seja uma superfície s dada por $f(x_1, x_2, \dots, x_n) = c$, de dimensão $(n-1)$ no $\mathbb{R}^n$. O plano tangente em $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ é o conjunto de todos os vetores tangentes às curvas sobre s que passam por x^* , ou seja, se $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ é tal que $f(\vec{x}(t)) = f(x_1(t), x_2(t), \dots, x_n(t)) = c$ e $\vec{x}(t^*) = (x_1(t^*), x_2(t^*), \dots, x_n(t^*)) = x^*$, então $\vec{x}'(t^*) = (x_1'(t^*), x_2'(t^*), \dots, x_n'(t^*))$ é um vetor tangente em x^* . Podemos nos perguntar como sabemos que a reunião desses vetores forma um “plano” ou, mais precisamente, como sabemos que o conjunto desses vetores constitui um subespaço de dimensão $(n-1)$ no $\mathbb{R}^n$. A resposta é simples: nem sempre o conjunto desses vetores forma um subespaço vetorial! As funções, tais que em qualquer ponto temos planos tangentes, são ditas diferenciáveis.

Pela regra da cadeia,

$$\begin{aligned} \frac{d}{dt} f(\vec{x}(t)) &= f(x_1(t), x_2(t), \dots, x_n(t)) = \\ \frac{\partial}{\partial x_1} f(x_1(t), x_2(t), \dots, x_n(t)) \cdot x_1'(t) + \dots + \frac{\partial}{\partial x_n} f(x_1(t), x_2(t), \dots, x_n(t)) \cdot x_n'(t) &= \\ \frac{\partial}{\partial x_1} f(\vec{x}(t)) \cdot x_1'(t) + \dots + \frac{\partial}{\partial x_n} f(\vec{x}(t)) \cdot x_n'(t). \end{aligned}$$

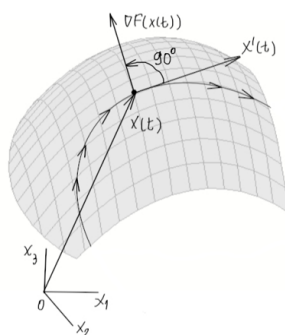
A expressão acima representa o produto escalar dos dois vetores

$$\text{grad}(f)(\vec{x}(t)) = \nabla f(\vec{x}(t)) = \left(\frac{\partial f}{\partial x_1}(\vec{x}(t)), \dots, \frac{\partial f}{\partial x_n}(\vec{x}(t)) \right),$$

que chamamos de vetor gradiente de f em $\vec{x}(t)$, ou simplesmente gradiente, e

$$\vec{x}'(t) = (x_1'(t), \dots, x_n'(t))$$

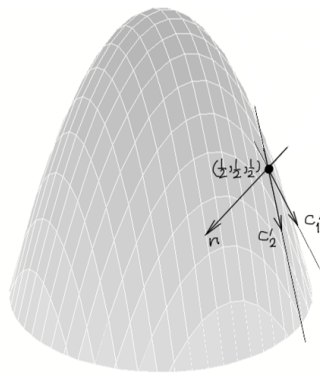
que é o vetor tangente em $\vec{x}(t)$. Numa notação mais compacta, $\frac{d}{dt}f(\vec{x}(t)) = \nabla f(\vec{x}(t)) \cdot \vec{x}'(t)$. Se a curva $\vec{x}(t)$ está sobre a superfície, então $f(\vec{x}(t)) = c$ e $\frac{d}{dt}f(\vec{x}(t)) = \nabla f(\vec{x}(t)) \cdot \vec{x}'(t) = 0$, portanto, $\nabla f(\vec{x}(t))$ e $\vec{x}'(t)$ são perpendiculares. Em outras palavras, o gradiente de f é perpendicular a todo vetor tangente $\vec{x}'(t)$ e, desse modo, ao plano tangente à superfície s em $\vec{x}(t)$. No gráfico abaixo, expressamos esse fato para uma superfície de dimensão 2 no $\mathbb{R}^3$:



Logo, é fácil escrever uma equação para o plano tangente no ponto x^* :

$$\frac{\partial f}{\partial x_1}(x^*) \cdot (x_1 - x_1^*) + \dots + \frac{\partial f}{\partial x_n}(x^*) \cdot (x_n - x_n^*) = 0.$$

Exemplo 4: No caso do parabolóide $f(x, y) = 1 - x^2 - y^2$, escrevemos a superfície como $z = 1 - x^2 - y^2$ ou $F(x, y, z) = 1 - x^2 - y^2 - z = 0$, que é a isoquanta de nível 0 de F . O gradiente é $\nabla F(x, y, z) = (-2x, -2y, -1)$. No ponto $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, resulta $n = \nabla F = (-1, -1, -1)$. Portanto, o plano tangente é $(-1) \cdot \left(x - \frac{1}{2}\right) + (-1) \cdot \left(y - \frac{1}{2}\right) + (-1) \cdot \left(z - \frac{1}{2}\right) = 0$ ou $x + y + z = \frac{3}{2}$. A reta perpendicular à superfície nesse ponto: $\left(\frac{1}{2} - r, \frac{1}{2} - r, \frac{1}{2} - r\right)$, que está desenhada no penúltimo gráfico. Como os vetores $c_1\left(\frac{1}{2}\right) = (1, 0, -1)$ e $c_2\left(\frac{1}{2}\right) = (0, 1, -1)$ são linearmente independentes, então eles geram o subespaço de dimensão 2 que tangencia a superfície no ponto $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. O gradiente e os dois vetores tangentes estão no gráfico abaixo:



Dada a função $f: \mathbb{R}^n \rightarrow \mathbb{R}$, então o plano tangente ao gráfico $(x_1, x_2, \dots, x_n, f(x_1, x_2, \dots, x_n)) = (x, f(x))$ de f no ponto

$$(x_1^*, x_2^*, \dots, x_n^*, f(x_1^*, x_2^*, \dots, x_n^*)) = (x^*, f(x^*))$$

é dado por:

$$z = f(x^*) + \nabla f(x^*) \cdot (x - x^*).$$

Pois, nesse caso, podemos apresentar o gráfico dessa função como a isoquanta de nível zero de $F: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$,

$$F(x_1, x_2, \dots, x_n, z) = F(x, z) = f(x_1, x_2, \dots, x_n) - z = f(x) - z.$$

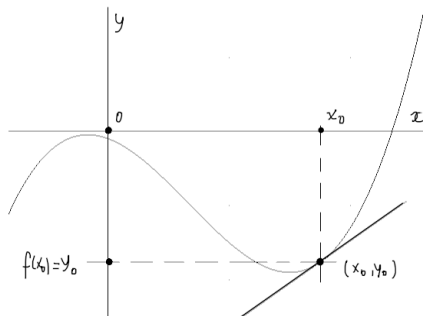
O gradiente de F , $\nabla F(x^*, z) = \left(\frac{\partial f}{\partial x_1}(x^*), \dots, \frac{\partial f}{\partial x_n}(x^*), -1 \right)$, é perpendicular à isoquanta $F(x, z) = 0$. Portanto, o plano tangente em (x^*, z) é $\nabla F(x^*, z) \cdot (x - x^*) = 0$, isto é, $\frac{\partial f}{\partial x_1}(x^*) \cdot (x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x^*) \cdot (x_2 - x_2^*) + \dots + \frac{\partial f}{\partial x_n}(x^*) \cdot (x_n - x_n^*) + (-1) \cdot (z - z^*) = 0$, em que $f(x^*) - z^* = 0$. Ou, ainda, $z = f(x_1^*, x_2^*, \dots, x_n^*) + \frac{\partial f}{\partial x_1}(x^*) \cdot (x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x^*) \cdot (x_2 - x_2^*) + \dots + \frac{\partial f}{\partial x_n}(x^*) \cdot (x_n - x_n^*)$.

Observe que a expressão acima, para funções $f(x) = y$ de uma variável x , resulta na usual equação da reta tangente ao gráfico de $(x^*, f(x^*))$:

$$y = f(x^*) + f'(x^*) \cdot (x - x^*).$$

O plano tangente ao gráfico da função $f(x) = y$ é uma generalização da reta tangente ao gráfico de uma função no plano $\mathbb{R}^2$. A curva $F(x, y) = f(x) - y = 0$ é o gráfico da função. O vetor $\nabla F(x_0, y_0) = (f'(x_0), -1)$ é perpendicular

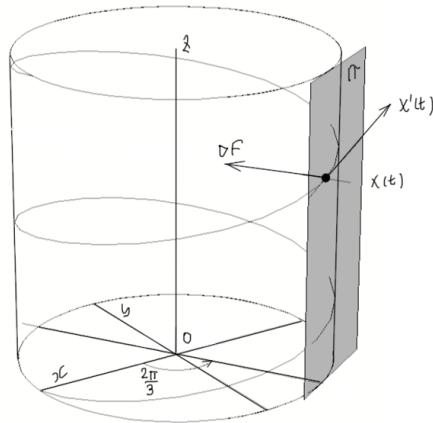
à reta tangente ao gráfico da função no ponto $P(x_0, y_0)$. Portanto, a reta tangente nesse ponto é $f'(x_0) \cdot (x - x_0) + (-1) \cdot (y - y_0) = 0$, isto é, $y = y_0 + f'(x_0) \cdot (x - x_0) = f(x_0) + f'(x_0) \cdot (x - x_0)$. Veja o desenho:



Exemplo 5: No desenho acima, $f(x) = x^3 - 4 \cdot x^2 - 3 \cdot x - 1$, se $P(x_0, y_0) = P\left(\frac{7}{2}, -\frac{141}{8}\right)$, a reta tangente ao gráfico neste ponto é $y = \frac{23}{4} \cdot \left(x - \frac{7}{2}\right) - \frac{141}{8}$.

Mostramos que todo vetor tangente à superfície em x^* é perpendicular ao $\nabla f(x^*)$. Um problema que resta para ser considerado é se todo o vetor perpendicular ao gradiente $\nabla f(x^*)$ é vetor tangente à superfície $f(x) = c$ em x^* . Voltaremos a esse problema mais à frente.

Exemplo 6: Consideremos o cilindro $F(x, y) = x^2 + y^2 = 1$. Em cada ponto sobre esse cilindro o vetor $\nabla F(x, y) = (2 \cdot x, 2 \cdot y, 0)$ é perpendicular às tangentes a esses pontos. Consideremos a hélice circular $c(t) = (\cos(t), \sin(t), t)$ sobre o cilindro, as tangentes $c'(t) = (-\sin(t), \cos(t), 1)$ tangenciam o cilindro e são perpendiculares aos vetores $\nabla F(c(t)) = (2 \cdot \cos(t), 2 \cdot \sin(t), 0)$, como é fácil verificar. No desenho abaixo, estão representados $c\left(\frac{8 \cdot \pi}{3}\right) = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}, \frac{8 \cdot \pi}{3}\right)$, $c'\left(\frac{8 \cdot \pi}{3}\right) = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, 1\right)$ e $\nabla F\left(c\left(\frac{8 \cdot \pi}{3}\right)\right) = (-1, \sqrt{3}, 0)$. O plano π tangente ao cilindro no ponto $c\left(\frac{8 \cdot \pi}{3}\right)$ é $-1 \cdot \left(x + \frac{1}{2}\right) + \sqrt{3} \cdot \left(y - \frac{\sqrt{3}}{2}\right) + 0 \cdot \left(z - \frac{8 \cdot \pi}{3}\right) = 0$:

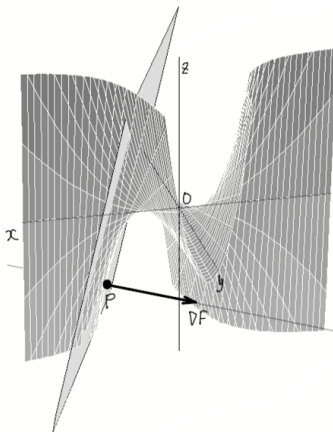


Em qualquer ponto (x_0, y_0, z_0) sobre o cilindro, o plano tangente nesse ponto é

$$2 \cdot x_0 \cdot (x - x_0) + 2 \cdot y_0 \cdot (y - y_0) + 0 \cdot (z - z_0) = 0, \text{ ou seja, } x_0 \cdot x + y_0 \cdot y = 1.$$

Exemplo 7: Considere a superfície que corresponde ao gráfico da função $f(x, y) = 3 \cdot x^2 \cdot y - x \cdot y$. Determine o plano que tangencia essa superfície no ponto $f(1, -1, f(1, -1))$.

A superfície determinada pelo gráfico da função pode ser representada pela equação $z = 3 \cdot x^2 \cdot y - x \cdot y$ ou pela isoquanta da função $F(x, y, z) = 3 \cdot x^2 \cdot y - x \cdot y - z = 0$. Em cada ponto $(x_0, y_0, z_0) = (x_0, y_0, f(x_0, y_0))$ sobre a superfície, o gradiente de F é o vetor $\nabla F(x_0, y_0, z_0) = (6 \cdot x_0 \cdot y_0 - y_0, 3 \cdot x_0^2 - x_0, -1)$. No ponto $P(1, -1, f(1, -1)) = P(1, -1, -2)$, o gradiente é $\nabla F(1, -1, -2) = (-5, 2, -1)$. O plano tangente à superfície nesse ponto é $5 \cdot x - 2 \cdot y + z = 5$. Observe o desenho abaixo:



Em geral, o gráfico no $\mathbb{R}^3$ de uma função $f(x,y)=z$ pode ser representado pela isoquanta $F(x,y,z)=f(x,y)-z=0$. O gradiente $\nabla F(x_0,y_0,z_0)=\left(\frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial y}(x_0,y_0), -1\right)$ é perpendicular ao plano tangente no gráfico da função $f(x,y)=z$ no ponto (x_0,y_0,z_0) . Portanto, o plano tangente nesse ponto é $\frac{\partial f}{\partial x}(x_0,y_0) \cdot (x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0) \cdot (y-y_0) + (-1) \cdot (z-z_0) = 0$ ou

$$z = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0) \cdot (x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0) \cdot (y-y_0),$$

em que $f(x_0,y_0)=z_0$.

Em geral, se $f(x_1,x_2,\dots,x_n)=y$, então o gráfico da função está representado pela isoquanta $F(x_1,x_2,\dots,x_n,y)=f(x_1,x_2,\dots,x_n)-y=0$, $\nabla F=\left(\frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n},-1\right)$. Daí concluímos que o plano que tangencia o gráfico no ponto $P(x_1^*,x_2^*,\dots,x_n^*,y^*)$ é

$$y = f(x_1^*,x_2^*,\dots,x_n^*) + \frac{\partial f}{\partial x_1}(x_1^*,x_2^*,\dots,x_n^*) \cdot (x_1-x_1^*) + \dots + \frac{\partial f}{\partial x_n}(x_1^*,x_2^*,\dots,x_n^*) \cdot (x_n-x_n^*),$$

que representamos como

$$y = y^* + \nabla f(x^*) \cdot (x - x^*),$$

em que $\nabla f(x^*)=\left(\frac{\partial f}{\partial x_1}(x^*),\dots,\frac{\partial f}{\partial x_n}(x^*)\right)$ e $x-x^*=\begin{pmatrix} x_1-x_1^* \\ \vdots \\ x_n-x_n^* \end{pmatrix}$. A última expressão para o plano tangente ao gráfico de uma função é essencial na lição sobre funções convexas e côncavas.

Nesta notação mais condensada, se $F(x_1,x_2,\dots,x_n)=c$ é uma superfície, então o plano tangente a essa superfície no ponto $x^*=(x_1^*,x_2^*,\dots,x_n^*)$ é

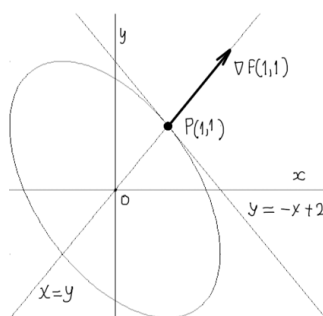
$$\nabla F(x^*) \cdot (x - x^*) = 0,$$

ou seja, o conjunto de vetores $x-x^*$ ortogonais ao $\nabla F(x^*)$.

Exemplo 8: Observe que o gráfico de $w=f(x,y,z)=x \cdot z - 2 \cdot x^2 - 2 \cdot y^2 - z^2 - 3 \cdot x - z$ está no $\mathbb{R}^4$ e é a isoquanta de $F(x,y,z,w)=x \cdot z - 2 \cdot x^2 - 2 \cdot y^2 - z^2 - 3 \cdot x - z - w=0$ de F . O hiperplano tangente ao gráfico de f no ponto $(1,1,-1,f(1,1,-1))$ é perpendicular ao vetor

$\nabla F(1,1,-8)=(-8,-4,2,-1)$, ou seja, é dado por $(-8)\cdot(x-1)+(-4)\cdot(y-1)+(2)\cdot(z+1)+(-1)\cdot(w+8)=0$ ou, ainda, $w=-8+(-8)\cdot(x-1)+(-4)\cdot(y-1)+(2)\cdot(z+1)$ e $8\cdot x+4\cdot y-2\cdot z+w=6$.

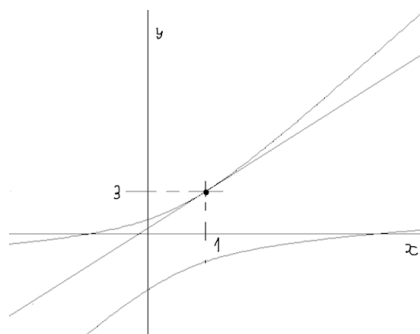
Exemplo 9: Seja a elipse $x^2+x\cdot y+y^2=3$, que é a isoquanta de nível três da função $F(x,y)=x^2+x\cdot y+y^2=3$, cujo gradiente é $\nabla F(x,y)=\begin{pmatrix} 2\cdot x+y \\ x+2\cdot y \end{pmatrix}$. $\nabla F(1,1)=\begin{pmatrix} 3 \\ 3 \end{pmatrix}$, isso significa que o vetor $n=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ é ortogonal à reta tangente à elipse no ponto $P(1,1)$. Portanto, a equação da reta tangente à elipse é $1\cdot(x-1)+1\cdot(y-1)=0$, isto é, $x+y=2$. A reta ortogonal na direção do vetor $n=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ é $x=1+s$, $y=1+s$, isto é, $y=x$ ou $x-y=0$.



Exercícios 1

- i) Achar a equação do plano tangente à superfície $z^3+x\cdot y\cdot z+y^2+x-1=0$ no ponto $P(1,0,1)$. Resposta: $x+y+3\cdot z=4$.
- ii) Achar a equação do plano tangente à superfície $x^2+2\cdot y^2=10-4\cdot z^2$ no ponto $P(2,-1,1)$. Resposta: $x-y+2\cdot z=5$.
- iii) Obtenha o plano tangente ao gráfico de $f(x,y)=\ln(x^2+y-1)$ no ponto $(\sqrt{e},1)$. Resposta: $z=1+\frac{2}{\sqrt{e}}\cdot(x-\sqrt{e})+\frac{1}{e}\cdot(y-1)$.
- iv) Obtenha o plano tangente ao gráfico de $f(x,y,z)=(x^2+y^2+z^2-1)^{1/2}$ no ponto $P(4,3,1)$. Resposta: $w=5+\frac{4}{5}\cdot(x-4)+\frac{3}{5}\cdot(y-3)+\frac{1}{5}\cdot(z-1)$.
- v) Obtenha a reta que tangencia a curva $x^3-y^3+x^2+3\cdot y^2-x=0$ no ponto $P(0,0)$.
- vi) Obtenha as retas que tangenciam a curva $x^2+y^2+x^2\cdot y=2$ nos pontos $P\left(1,\frac{-1\pm\sqrt{5}}{2}\right)$.
- vii) Obtenha a reta tangente ao gráfico da hipérbole $x^2-4x\cdot y+y^2-3\cdot x+3\cdot y-4=0$ no ponto $P(1,3)$. Obtenha também a equação da reta perpendicular à

tangente nesse ponto.



3. Superfícies de dimensão $(n-k)$ em $\mathbb{R}^n$

Como observamos anteriormente, a reta $\left(\frac{3}{2}-t, -\frac{1}{4}, t\right)$ é definida pela interseção dos planos $\pi_1: 2 \cdot x + y + z = 1$ e $\pi_2: x - 2 \cdot y + z = 2$. Para nós, essa reta é uma “superfície” de dimensão 1 no $\mathbb{R}^3$.

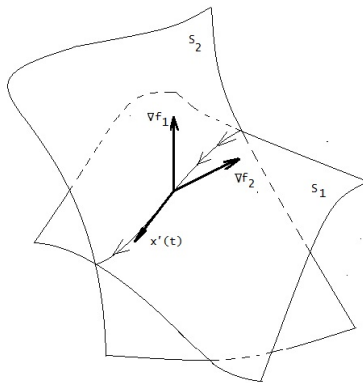
Sob condições que explicitamos mais abaixo, a solução de um sistema de equações

$$\begin{aligned} S_1: f_1(x_1, x_2, \dots, x_n) &= c_1, \\ S_2: f_2(x_1, x_2, \dots, x_n) &= c_2, \\ &\vdots \\ S_k: f_k(x_1, x_2, \dots, x_n) &= c_k, \end{aligned}$$

é uma superfície $S = S_1 \cap S_2 \cap \dots \cap S_k$ de dimensão $n-k$ no $\mathbb{R}^n$. Se $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ é uma curva sobre essa superfície S , então $f_i(x(t)) = f_i(x_1(t), x_2(t), \dots, x_n(t)) = c_i$, para $1 \leq i \leq k$, logo

$$\frac{d}{dt} f_i(\vec{x}(t)) = \nabla f_i(\vec{x}(t)) \cdot \vec{x}'(t) = 0.$$

Em outras palavras, $\vec{x}'(t)$ é ortogonal aos vetores $\nabla f_1(\vec{x}(t)), \dots, \nabla f_k(\vec{x}(t))$, ou ainda, $\vec{x}'(t)$ é ortogonal ao subespaço vetorial gerado pelos vetores $\nabla f_i(\vec{x}(t))$, $1 \leq i \leq k$.



A condição que vamos exigir para que s defina uma superfície de **dimensão** $n-k$ é que os vetores $\nabla f_i(\vec{x}(t))$ e $1 \leq i \leq k$ sejam linearmente independentes, pois, nesse caso, o subespaço gerado pelos vetores tangentes $\vec{x}'(t)$, ou seja, o “plano” tangente em $x(t)$, tem dimensão $n-k$. Formulamos essa condição de maneira mais algébrica, exigindo que o posto da matriz cujas linhas são os gradientes $\nabla f_i(\vec{x}(t))$ seja k , isto é,

$$\text{posto} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_k}{\partial x_1}(x) & \frac{\partial f_k}{\partial x_2}(x) & \dots & \frac{\partial f_k}{\partial x_n}(x) \end{pmatrix} = k.$$

Em tempo, o posto de uma matriz $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$ é a ordem maior das submatrizes de A com determinante não nulo.

Exemplo 8: $f(x, y, z) = x^2 + y^2 + 6 \cdot z^2 = 5$ define uma superfície em $\mathbb{R}^3$ de dimensão 2, pois $\nabla f(x, y, z) = (2 \cdot x, 2 \cdot y, 12 \cdot z) \neq \vec{0}$ em todo o ponto (x, y, z) sobre essa superfície.

Exemplo 9:

$$x_1 + 2 \cdot x_2 - x_3 + 3 \cdot x_4 = 5$$

$$3 \cdot x_1 - x_2 + 4 \cdot x_3 - x_4 = 1$$

define uma superfície de dimensão 2 em $\mathbb{R}^4$, pois se $A = \begin{pmatrix} 1 & 2 & -1 & 3 \\ 3 & -1 & 4 & -1 \end{pmatrix}$, verificamos que $\text{posto } A = 2$. Por linha-redução:

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 3 & 5 \\ 3 & -1 & 4 & -1 & 1 \end{array} \right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 1 & \frac{1}{7} & 5 \\ 0 & 1 & -1 & \frac{10}{7} & 2 \end{array} \right),$$

portanto, a superfície de dimensão 2, que chamamos de uma variedade linear, é dada por

$$x_1 = 5 - x_3 - \frac{1}{7} \cdot x_4,$$

$$x_2 = 2 + x_3 - \frac{10}{7} \cdot x_4.$$

A dimensão é 2 porque temos duas variáveis independentes nas equações acima, que, em outras palavras, significa dizer que os gradientes $\vec{n}_1 = (1, 2, -1, 3)$ e $\vec{n}_2 = (3, -1, 4, -1)$ são linearmente independentes.

Exemplo 10:

$$2 \cdot x + y + z = 1$$

$$x + 2 \cdot y + z = 2$$

define uma superfície de dimensão 1 em $\mathbb{R}^3$ pois

$$\text{posto} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = 2.$$

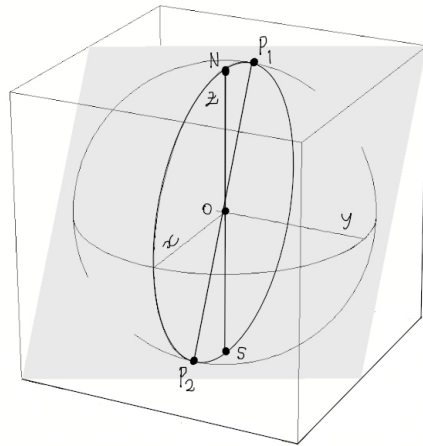
A “superfície” acima é a reta que está na interseção dos dois planos. O posto igual a dois da matriz acima expressa que os dois gradientes, $\vec{n}_1 = (2, 1, 1)$ e $\vec{n}_2 = (1, 2, 1)$, são linearmente independentes. Logo, os planos não são paralelos e se interceptam na reta $x = -\frac{1}{3} \cdot t$, $y = 1 - \frac{1}{3} \cdot t$, $z = t$, que se obtém por linha-redução da matriz do sistema.

Exemplo 11: Seja a “superfície” dada pela interseção da esfera e do plano abaixo:

$$x^2 + y^2 + z^2 = 25$$

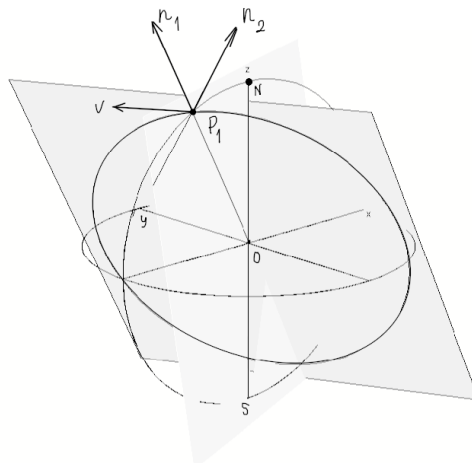
$$2 \cdot x + y + z = 0.$$

Obviamente, o sistema acima define uma “superfície” de dimensão 1, que é uma circunferência centrada na origem de raio $r=5$. No gráfico abaixo, tal circunferência é visível:



Os pontos $P_1(-\sqrt{5}, 0, 2\sqrt{5})$ e $P_2(\sqrt{5}, 0, -2\sqrt{5})$ estão sobre a circunferência que está na interseção da esfera e do plano. Os pontos $N(5, 0, 0)$ e $S(-5, 0, 0)$ são os polos da esfera.

Os gradientes da esfera e do plano são $\nabla g_1 = (2 \cdot x, 2 \cdot y, 2 \cdot z)$ e $\nabla g_2 = (2, 1, 1)$, respectivamente. No ponto $P_1(-\sqrt{5}, 0, 2\sqrt{5})$, $\nabla g_1 = n_1 = (-2\sqrt{5}, 0, 4\sqrt{5})$ e $\nabla g_2 = n_2 = (2, 1, 1)$ estão num plano perpendicular ao plano $2 \cdot x + y + z = 0$. O vetor v , que tangencia a circunferência definida por $x^2 + y^2 + z^2 = 25$ e $2 \cdot x + y + z = 0$, é ortogonal ao n_1 e ao n_2 e está no plano $2 \cdot x + y + z = 0$, ou seja, v é ortogonal ao n_1 e ao n_2 .



Embora, tudo pareça muito simples, nem tudo foi totalmente esclarecido. Por exemplo, mostramos que todo vetor que tangencia a superfície é perpendicular ao conjunto de k gradientes das isoquantas que definem a superfície, portanto, esses vetores estão no subespaço ortogonal a esses k gradientes. Esse subespaço ortogonal tem dimensão $n-k$. Mas, não mostramos que todo vetor que é ortogonal aos gradientes é um vetor tangente! Em outras palavras, falta mostrar que a dimensão do plano tangente é $n-k$. Falta também relacionar a dimensão do plano tangente ao número de variáveis independentes e dependentes que temos em $x_1, x_2, \dots, x_n$. Vamos admitir que essas proposições sejam verdadeiras.

4. Superfícies tangentes

Observe que a esfera $(x-1)^2 + (y-1)^2 + (z-1)^2 = 12$ e o plano $2 \cdot x + 2 \cdot y + 2 \cdot z = 18$ se interceptam no ponto $P(3,3,3)$. Entretanto, não é evidente que se tangenciem nesse ponto. Como podemos verificar isso? Duas superfícies que se tangenciam num certo ponto $P(x^*, y^*, z^*)$ têm nesse ponto o mesmo plano tangente. Em relação à esfera, os vetores perpendiculares aos planos tangentes num ponto $P(x, y, z)$ são dados por $\vec{n}_1 = (2 \cdot (x-1), 2 \cdot (y-1), 2 \cdot (z-1))$. Portanto, em $P(3,3,3)$, $\vec{n}_1 = (4, 4, 4)$, o plano tangente à esfera neste ponto é $4 \cdot (x-3) + 4 \cdot (y-3) + 4 \cdot (z-3) = 0$ ou $x + y + z = 9$, que é o plano acima. Se duas

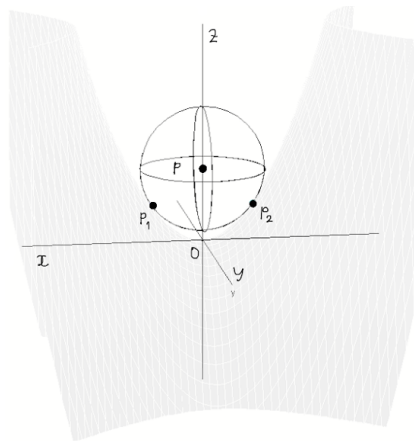
superfícies definidas por $f_1(x, y, z) = c_1$ e $f_2(x, y, z) = c_2$ se interceptam num ponto $P(x^*, y^*, z^*)$ e têm nesse ponto o mesmo plano tangente, então os gradientes são paralelos, isto é, existe $\lambda^* \neq 0$ tal que $\nabla f_2(x^*, y^*, z^*) = \lambda^* \cdot \nabla f_1(x^*, y^*, z^*)$. Como no exemplo acima, o plano $2 \cdot x + 2 \cdot y + 2 \cdot z = 18$ é o plano que tangencia a esfera.

Duas superfícies definidas por $f_1(x_1, x_2, \dots, x_n) = c_1$ e $f_2(x_1, x_2, \dots, x_n) = c_2$, que se interceptam no ponto $P(x_1^*, x_2^*, \dots, x_n^*) = P(x^*)$, se tangenciam nesse ponto se, e só se, existe $\lambda^* \neq 0$ tal que $\nabla f_2(x_1^*, x_2^*, \dots, x_n^*) = \lambda^* \cdot \nabla f_1(x_1^*, x_2^*, \dots, x_n^*)$. Nesse caso, obviamente, o plano tangente comum é $\nabla f_2(x^*) \cdot (x - x^*) = \nabla f_1(x^*) \cdot (x - x^*) = 0$.

Exemplo 12: Determine as esferas $x^2 + y^2 + (z-1)^2 = r^2$, centradas em $P(0, 0, 1)$, que tangenciam o elipsoide $z = x^2 - y^2$. Estamos procurando os pontos $P(x, y, z)$, tais que

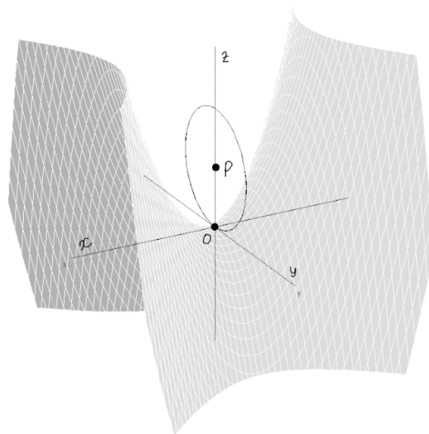
$$(2 \cdot x, 2 \cdot y, 2 \cdot (z-1)) = \lambda \cdot (2 \cdot x, -2 \cdot y, -1)$$

para algum $\lambda \neq 0$ e $z = x^2 - y^2$. Se $x \neq 0$, então $\lambda = 1$, $y = 0$, $z = \frac{1}{2}$ e $x = \pm \frac{\sqrt{2}}{2}$. Os pontos $P_1\left(\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ e $P_2\left(-\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ estão sobre a esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$, logo, essa esfera tangencia o elipsoide nesses pontos.



Se $y \neq 0$, então $\lambda = -1$, $x = 0$, $z = \frac{3}{2}$ e $\frac{3}{2} = -y^2$, o que é impossível. Se $x = 0$ e $y = 0$, então $z = 0$. Nesse caso, os gradientes acima se reduzem a $(0, 0, -2) = 2 \cdot (0, 0, -1)$, ou seja,

$\lambda = 2$. Portanto, a origem $(0,0,0)$ está sobre a esfera $x^2 + y^2 + (z-1)^2 = 1$, que tangencia o elipsoide nesse ponto.



Exercício 2

- i) Determine o ponto onde a esfera $x^2 + y^2 + z^2 = 6$ e o plano $x - y + 2 \cdot z = 6$ se tangenciam.
- ii) Determine o ponto onde a esfera $x^2 + y^2 + z^2 = r^2$, centrada na origem, e o plano $3 \cdot x - 2 \cdot y + z = 4$ se tangenciam.
- iii) Encontre o raio r da esfera $x^2 + (y-1)^2 + z^2 = r^2$ que tangencia o plano $2 \cdot x + y - z = 4$.
- iv) Mostre que as superfícies $x^2 + y^2 + z^2 = 2$ e $x \cdot y = 1$ se tangenciam no ponto $(1,1,0)$.
- v) Determine a isoquanta $c = x^{\frac{1}{3}} \cdot y^{\frac{1}{3}} \cdot z^{\frac{1}{3}}$ que tangencia o plano $x + y + z = 90$.
- vi) Determine os números c tal que o plano $x + 2 \cdot y + z = c$ tangencie o elipsoide $x^2 + 4 \cdot y^2 + z^2 = 1$.

Em muitas situações, as superfícies que estão se tangenciando não são da mesma dimensão. No exemplo anterior, temos circunferências tangenciando o hiperboloide. Dizemos que uma superfície de dimensão menor tangencia uma superfície de dimensão maior em determinado ponto P se o plano tangente à primeira superfície está contido no plano tangente à segunda superfície em P .

5. Lagrangeanos

Exemplo 13: Queremos determinar as esferas $f(x, y, z) = x^2 + y^2 + (z-1)^2 = r^2$ centradas no ponto $P(0, 0, 1)$ que tangenciam o hiperboloide $g(x, y, z) = x^2 - y^2 - z = c = 0$. De fato, queremos determinar quais os raios das esferas e os pontos de contato. Na verdade, resolvemos esse problema no Exemplo 11. Já sabemos que as esferas $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$ e $x^2 + y^2 + (z-1)^2 = 1$ são tangenciais à sela. A esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$ tangencia a sela nos pontos $P_1\left(\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ e $P_2\left(-\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ e, por sua vez, a esfera $x^2 + y^2 + (z-1)^2 = 1$ na origem $(0, 0, 0)$.

Duas superfícies se tangenciam num ponto $P(x^*, y^*, z^*)$ se nesse ponto têm o mesmo plano tangente. Portanto, a solução do nosso problema é expressa pelas equações $\nabla f(x^*, y^*, z^*) = \lambda^* \cdot \nabla g(x^*, y^*, z^*)$ e $g(x^*, y^*, z^*) = x^{*2} - y^{*2} - z^* = 0$. A primeira das equações expressa a direção perpendicular ao plano tangente comum, e a segunda equação diz que o ponto comum $P(x^*, y^*, z^*)$ está sobre o hiperboloide. Escrevendo novamente essas equações temos: $(2 \cdot x^*, 2 \cdot y^*, 2 \cdot (z^* - 1)) = \lambda^* \cdot (2 \cdot x^*, -2 \cdot y^*, -1)$ e $x^{*2} - y^{*2} - z^* = 0$ ou, ainda,

$$\begin{aligned} 2 \cdot x^* &= \lambda^* \cdot 2 \cdot x^* \\ 2 \cdot y^* &= -\lambda^* \cdot 2 \cdot y^* \\ 2 \cdot (z^* - 1) &= -\lambda^* \\ x^{*2} - y^{*2} - z^* &= 0 \end{aligned}$$

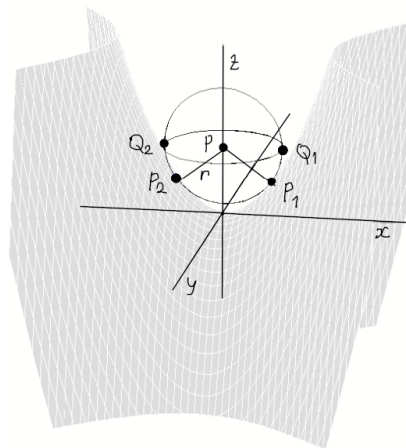
As condições acima $\nabla f(x^*, y^*, z^*) = \lambda^* \cdot \nabla g(x^*, y^*, z^*)$ e $x^{*2} - y^{*2} - z^* = 0$ expressamos pelas condições de primeira ordem do lagrangeano $\mathcal{L}(x, y, z) = x^2 + y^2 + (z-1)^2 - \lambda \cdot (x^2 - y^2 - z)$. Se

$$\mathcal{L}(x, y, z) = f(x, y, z) - \lambda \cdot (g(x, y, z) - c),$$

então,

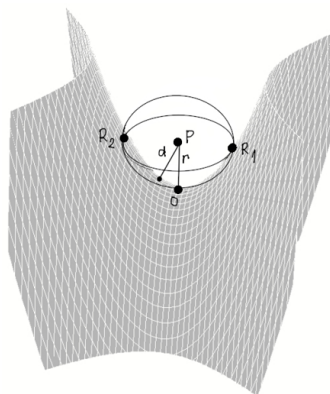
$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x}(x, y, z, \lambda) &= \frac{\partial f}{\partial x}(x, y, z) - \lambda \cdot \frac{\partial g}{\partial x}(x, y, z) = 0 \\
\frac{\partial \mathcal{L}}{\partial y}(x, y, z, \lambda) &= \frac{\partial f}{\partial y}(x, y, z) - \lambda \cdot \frac{\partial g}{\partial y}(x, y, z) = 0 \\
\frac{\partial \mathcal{L}}{\partial z}(x, y, z, \lambda) &= \frac{\partial f}{\partial z}(x, y, z) - \lambda \cdot \frac{\partial g}{\partial z}(x, y, z) = 0' \\
\frac{\partial \mathcal{L}}{\partial \lambda}(x, y, z, \lambda) &= -\frac{\partial g}{\partial \lambda}(x, y, z) = 0
\end{aligned}$$

As incógnitas nesse problema são o ponto (x, y, z) e o multiplicador λ , portanto o que as condições de primeira ordem do lagrangeano expressam são as condições necessárias e suficientes para que uma isoquanta da função $f(x, y, z) = r$, no caso uma esfera $x^2 + y^2 + (z-1)^2 = r^2$, tangencie a superfície $g(x, y, z) = c$, no caso $x^2 - y^2 - z = 0$. No exemplo, encontramos duas esferas de raios $r = \frac{\sqrt{3}}{2}$ e $r = 1$. Se supomos $x^* \neq 0$, então $\lambda^* = 1$, $y^* = 0$ e $z^* = \frac{1}{2}$, que resulta nos pontos $P\left(\pm \frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ que estão sobre a esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$. Se supomos $y^* \neq 0$, então $\lambda^* = -1$, $x^* = 0$ e $z^* = \frac{3}{2}$, que resulta na equação sem solução $-y^2 = \frac{3}{2}$. Portanto, necessariamente $y^* = 0$. Se, por outro lado, supomos $x^* = 0$ e $y^* = 0$, então obtemos a solução $(0, 0, 0)$ com $\lambda^* = -2$, ou seja, a esfera $x^2 + y^2 + (z-1)^2 = 1$ tangencia o hiperboloide na origem. Os pontos $P_1\left(\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ e $P_2\left(-\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$, sobre o hiperboloide, são os pontos que estão mais próximos do ponto $P(0, 0, 1)$. Essa distância é $r = \frac{\sqrt{3}}{2} < 1$, que corresponde ao raio da esfera. Os outros pontos no desenho são $Q_1\left(\frac{\sqrt{3}}{2}, 0, 1\right)$ e $Q_2\left(-\frac{\sqrt{3}}{2}, 0, 1\right)$.

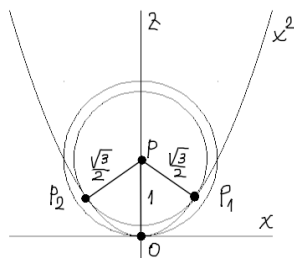


A outra esfera que tangencia a sela é $x^2 + y^2 + (z-1)^2 = 1 = r$, cujo ponto de tangência está na origem $(0, 0, 0)$. Os pontos $R_1(1, 0, 1)$ e $R_2(-1, 0, 1)$ são aqueles

onde a esfera encontra a superfície $x^2 - y^2 = z$. Existem pontos sobre $x^2 - y^2 = z$ que estão à distância $d < 1$ dentro da esfera $x^2 + y^2 + (z-1)^2 = 1 = r^2$.



Os pontos $P_1\left(\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ e $P_2\left(-\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ estão no interior da esfera $x^2 + y^2 + (z-1)^2 = 1$. Veja o desenho:



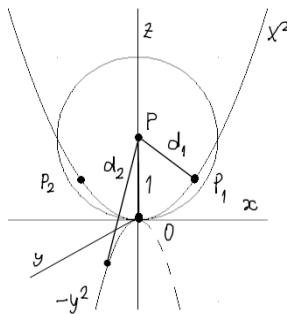
Poderíamos reformular nosso problema: quais os pontos sobre o hiperboloide que estão mais próximos do ponto $P(0,0,1)$? Em outras palavras, temos o problema:

$$\begin{aligned} \min & x^2 + y^2 + (z-1)^2 \\ \text{sa} & x^2 - y^2 - z = 0. \end{aligned}$$

A solução desse problema é dada por alguma das esferas $x^2 + y^2 + (z-1)^2 = r^2$ que tangenciam o hiperboloide! Já sabemos que os pontos de mínimo estão sobre a esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$. Se definimos o lagrangeano $\mathcal{L}(x, y, z, \lambda) = x^2 + y^2 + (z-1)^2 - \lambda \cdot (x^2 - y^2 - z)$, então as condições de primeira ordem $\frac{\partial \mathcal{L}}{\partial x} = 0$, $\frac{\partial \mathcal{L}}{\partial y} = 0$, $\frac{\partial \mathcal{L}}{\partial z} = 0$ e $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ expressam as condições de tangência tratadas acima.

A esfera $x^2 + y^2 + (z-1)^2 = 1$ satisfaz as condições de tangência no ponto $P(0,0,1)$, mas esse ponto não corresponde nem a um ponto de mínimo local e

nem a um ponto de máximo local de $f(x,y,z)=x^2+y^2+(z-1)^2$ sujeita à restrição $x^2-y^2-z=0$, como podemos verificar nos gráficos abaixo: $d_1 < 1 < d_2$.



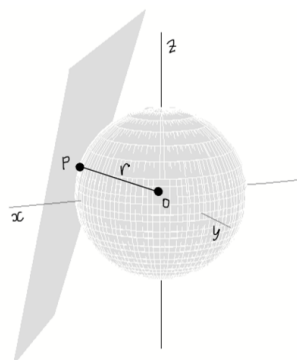
O lagrangeano não responde à pergunta sobre máximos ou mínimos, mas sobre isoquantas de f que tangenciam a restrição $g=0$. Por isso, as condições de primeira ordem do lagrangeano para mínimos e máximos são condições necessárias, mas não suficientes.

No gráfico acima, observamos que $(0,0,0)$ é uma sela para $f(x,y,z)=x^2+y^2+(z-1)^2$, pois os pontos $P(x,0,x^2)$ sobre o hiperboloide, próximos à origem, estão a distâncias $d < 1$; os pontos $P(0,y,-y^2)$ estão a distâncias $d > 1$; a origem dista 1 do ponto $P(0,0,1)$.

Exercício 3

No exercício 2, item ii), determinamos o ponto onde a esfera $x^2+y^2+z^2=r^2$ centrada na origem e o plano $3 \cdot x - 2 \cdot y + z = 4$ se tangenciam. Determine a solução do problema:

$$\begin{aligned} \min \quad & x^2 + y^2 + z^2 \\ \text{sa} \quad & 3 \cdot x - 2 \cdot y + z = 4. \end{aligned}$$



Resposta: $P\left(\frac{6}{7}, -\frac{4}{7}, \frac{2}{7}\right)$ e $r = \frac{2 \cdot \sqrt{14}}{7}$

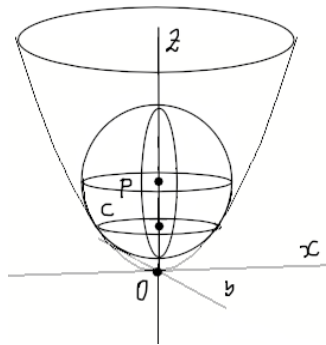
Exemplo 14: Seja o problema:

$$\begin{aligned} \min \quad & x^2 + y^2 + (z-1)^2 \\ \text{sa} \quad & x^2 + y^2 - z = 0. \end{aligned}$$

Observe que a restrição é o parabolóide $x^2 + y^2 - z = 0$. Nos pontos de mínimo, caso existam, existe uma esfera $x^2 + y^2 + (z-1)^2 = r^2$ que tangencia o parabolóide $x^2 + y^2 - z = 0$. Descobrimos as esferas que tangenciam o parabolóide pelas condições de primeira ordem do lagrangeano $\mathcal{L}(x, y, z, \lambda) = x^2 + y^2 + (z-1)^2 - \lambda \cdot (x^2 + y^2 - z)$:

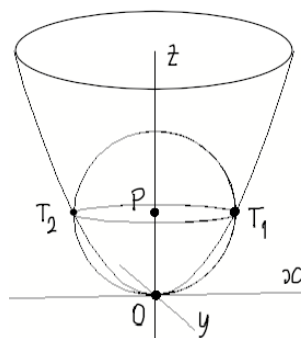
$$\frac{\partial \mathcal{L}}{\partial x}(x, y, z, \lambda) = 2 \cdot x \cdot (1 - \lambda) = 0, \quad \frac{\partial \mathcal{L}}{\partial y}(x, y, z, \lambda) = 2 \cdot y \cdot (1 - \lambda) = 0, \quad \frac{\partial \mathcal{L}}{\partial z}(x, y, z, \lambda) = 2 \cdot (z-1) + \lambda = 0 \text{ e } \frac{\partial \mathcal{L}}{\partial \lambda}(x, y, z, \lambda) = x^2 + y^2 - z = 0.$$

Se $\lambda = 1$, então a solução $x^2 + y^2 = 1, z = \frac{1}{2}$ é uma circunferência sobre a esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$. A circunferência $c: x^2 + y^2 = 1, z = \frac{1}{2}$ é constituída por pontos de tangência entre a esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{4}$ e o parabolóide $x^2 + y^2 - z = 0$.



O círculo $x^2 + y^2 = \frac{1}{2}, z = \frac{1}{2}$ é constituído de pontos de mínimos globais para $f(x, y, z) = x^2 + y^2 + (z-1)^2$, cuja distância entre o ponto $P(0, 0, 1)$ e o ponto $P(0, 0, 0)$ é $\frac{\sqrt{3}}{2}$. Entretanto, chegamos a essa conclusão analisando o desenho acima. O lagrangeano nos informa apenas sobre a tangência.

Se $\lambda \neq 1$, então $x = y = z = 0$, que corresponde à origem $(0, 0, 0)$, está sobre a esfera $x^2 + y^2 + (z-1)^2 = 1$. A esfera $x^2 + y^2 + (z-1)^2 = 1$ tangencia o parabolóide no ponto $(0, 0, 0)$.

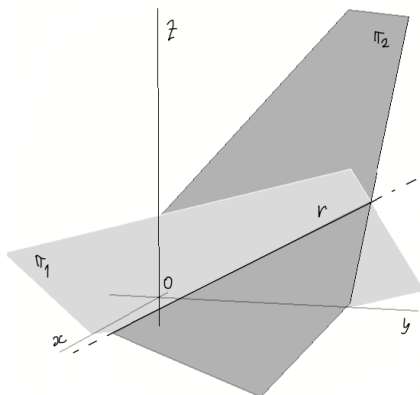


Os pontos $T_1(1,0,1)$ e $T_2(-1,0,1)$ são pontos sobre a circunferência $x^2 + y^2 = 1$ e $z = 1$. Essa circunferência é constituída por pontos onde a esfera $x^2 + y^2 + (z-1)^2 = 1$ corta o paraboloide. Observando o gráfico, concluímos que $(0,0,0)$ é um máximo local para $x^2 + y^2 + (z-1)^2$ sujeita à restrição $x^2 + y^2 = z$.

Exemplo 15: Consideremos a reta r definida na interseção dos planos

$$\begin{aligned}\pi_1 : x + y - 2 \cdot z &= 3 \\ \pi_2 : x - 2 \cdot y + z &= 2\end{aligned}$$

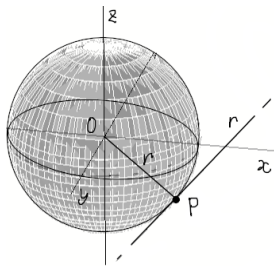
que, parametrizada por z , resulta $x = \frac{8}{3} + z, y = \frac{1}{3} + z$:



O problema é determinar o ponto sobre a reta r que está mais próximo da origem:

$$\begin{aligned}\min \quad & x^2 + y^2 + z^2 \\ \text{sa} \quad & x + y - 2 \cdot z = 3 \\ & x - 2 \cdot y + z = 2\end{aligned}$$

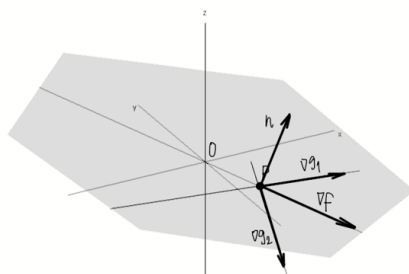
No ponto onde se encontra esse mínimo, uma esfera $f(x, y, z) = x^2 + y^2 + z^2 = R^2$ tangencia a reta $r: x = \frac{8}{3} + z, y = \frac{1}{3} + z$, ou seja, $\nabla f(x, y, z) = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ 2 \cdot z \end{pmatrix} = \lambda_1 \cdot \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \lambda_1 \cdot \nabla g_1(x, y, z) + \lambda_2 \cdot \nabla g_2(x, y, z)$. Logo, temos $x = \frac{\lambda_1 + \lambda_2}{2}$, $y = \frac{\lambda_1 - 2\lambda_2}{2}$ e $z = -\lambda_1 + \frac{\lambda_2}{2}$. Como $x + y - 2 \cdot z = 3$ e $x - 2 \cdot y + z = 2$, substituindo x , y e z , resulta $2 \cdot \lambda_1 - \lambda_2 = 2$ e $\lambda_1 - 2 \cdot \lambda_2 = -\frac{4}{3}$; $\lambda_1 = \frac{16}{9}$ e $\lambda_2 = \frac{14}{9}$; $x = \frac{5}{3}$, $y = -\frac{2}{3}$ e $z = -1$.



No ponto $P(\frac{5}{3}, -\frac{2}{3}, -1)$, o gradiente $\nabla f(\frac{5}{3}, -\frac{2}{3}, -1)$ é uma combinação linear dos gradientes $\nabla g_1(\frac{5}{3}, -\frac{2}{3}, -1)$ e $\nabla g_2(\frac{5}{3}, -\frac{2}{3}, -1)$:

$$\nabla f = \begin{pmatrix} \frac{10}{3} \\ -\frac{4}{3} \\ -2 \end{pmatrix} = \frac{16}{9} \cdot \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} + \frac{14}{9} \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \lambda_1 \cdot \nabla g_1 + \lambda_2 \cdot \nabla g_2.$$

Esses vetores são ortogonais ao vetor $n = (1, 1, 1)$, como facilmente se observa. O plano que contém o ponto $P(\frac{5}{3}, -\frac{2}{3}, -1)$ e é ortogonal ao vetor $n = (1, 1, 1)$ é o plano $x + y + z = 0$. Veja a figura abaixo:



O lagrangeano afirma algo simples: no ponto onde uma isoquanta de f tangencia a “superfície” $g_1 \cap g_2$, no caso uma reta, os gradientes ∇f , ∇g_1 e ∇g_2 estão num mesmo plano. Como ∇g_1 e ∇g_2 são linearmente independentes por hipótese, então ∇f é uma combinação linear de ambos.

Exercício 4

Determine a distância que a reta determinada por $x + y + z - 5 = 0$ e $x + 2 \cdot y + z - 2 = 0$ está da origem.

Exemplo 16: Dado o plano $\pi: a \cdot x + b \cdot y + c \cdot z + d = 0$, que não contém o ponto $P_0(x_0, y_0, z_0)$, determinar o ponto $P^*(x^*, y^*, z^*)$ sobre o plano π que está mais próximo de $P_0(x_0, y_0, z_0)$. A solução desse problema se resume a encontrar a esfera $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$, que tangencia o plano π . Essa condição geométrica pode ser expressa pelas condições de primeira ordem do lagrangeano:

$$\mathcal{L}(x, y, z, \lambda) = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 - \lambda \cdot (a \cdot x + b \cdot y + c \cdot z + d),$$

ou seja,

$$\frac{\partial \mathcal{L}}{\partial x} = 2 \cdot (x - x_0) - \lambda \cdot a = 0;$$

$$\frac{\partial \mathcal{L}}{\partial y} = 2 \cdot (y - y_0) - \lambda \cdot b = 0;$$

$$\frac{\partial \mathcal{L}}{\partial z} = 2 \cdot (z - z_0) - \lambda \cdot c = 0;$$

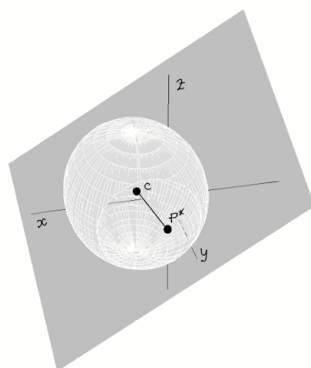
$$\frac{\partial \mathcal{L}}{\partial \lambda} = a \cdot x + b \cdot y + c \cdot z + d = 0.$$

Essas condições estão determinando um ponto $P^*(x^*, y^*, z^*)$ sobre o plano, de maneira que a esfera $(x^* - x_0)^2 + (y^* - y_0)^2 + (z^* - z_0)^2 = r^{*2}$ seja tangente ao plano nesse ponto. Pelas equações acima, $x^* = \frac{\lambda^* \cdot a}{2} + x_0$, $y^* = \frac{\lambda^* \cdot b}{2} + y_0$ e $z^* = \frac{\lambda^* \cdot c}{2} + z_0$, logo $0 = a \cdot x^* + b \cdot y^* + c \cdot z^* + d = \frac{\lambda^*}{2} \cdot (a^2 + b^2 + c^2) + a \cdot x_0 + b \cdot y_0 + c \cdot z_0 + d$. Portanto, $\lambda^* = -\frac{2 \cdot (a \cdot x_0 + b \cdot y_0 + c \cdot z_0 + d)}{(a^2 + b^2 + c^2)}$ (*).

Observe que podemos determinar x^* , y^* e z^* utilizando as expressões $x^* = \frac{\lambda^* \cdot a}{2} + x_0$, $y^* = \frac{\lambda^* \cdot b}{2} + y_0$ e $z^* = \frac{\lambda^* \cdot c}{2} + z_0$. Por outro lado, as equações acima: $2 \cdot (x^* - x_0) = \lambda^* \cdot a$, $2 \cdot (y^* - y_0) = \lambda^* \cdot b$ e $2 \cdot (z^* - z_0) = \lambda^* \cdot c$, implicam $\lambda^{*2} = \frac{4 \cdot [(x^* - x_0)^2 + (y^* - y_0)^2 + (z^* - z_0)^2]}{(a^2 + b^2 + c^2)} = \frac{4 \cdot r^{*2}}{(a^2 + b^2 + c^2)}$ (**).

Reunindo os dois resultados – (*) e (**) –, temos $\lambda^{*2} = \frac{4 \cdot (a \cdot x_0 + b \cdot y_0 + c \cdot z_0 + d)^2}{(a^2 + b^2 + c^2)^2} = \frac{4 \cdot r^{*2}}{(a^2 + b^2 + c^2)}$. Portanto, $r^{*2} = \frac{(a \cdot x_0 + b \cdot y_0 + c \cdot z_0 + d)^2}{(a^2 + b^2 + c^2)}$. Então, $r^* = \frac{|a \cdot x_0 + b \cdot y_0 + c \cdot z_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$ é a distância entre o plano e o ponto.

Exemplo 17: Considere o problema de determinar o ponto $P^*(x^*, y^*, z^*)$ sobre o plano $3 \cdot x + 2 \cdot y + 5 \cdot z = 1$ que está mais próximo do ponto $C(1, 1, 1)$. Geometricamente, esse problema se traduz em encontrar o ponto $P^*(x^*, y^*, z^*)$ no qual uma esfera $(x-1)^2 + (y-1)^2 + (z-1)^2 = r^2$ tangencia o plano. Veja o desenho.



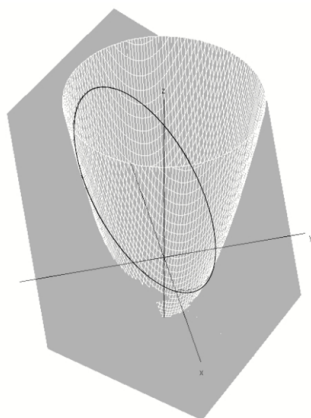
No ponto $P^*(x^*, y^*, z^*)$, o vetor normal à esfera e o vetor normal ao plano são paralelos, o que expressamos através das condições de primeira ordem do lagrangeano $\mathcal{L}(x, y, z, \lambda) = (x-1)^2 + (y-1)^2 + (z-1)^2 - \lambda \cdot (3 \cdot x + 2 \cdot y + 5 \cdot z - 1)$. Isto é,

$\frac{\partial \mathcal{L}}{\partial x} = 2 \cdot x - 3 \cdot \lambda - 2 = 0$, $\frac{\partial \mathcal{L}}{\partial y} = 2 \cdot y - \lambda - 2 = 0$, $\frac{\partial \mathcal{L}}{\partial z} = 2 \cdot z - 5 \cdot \lambda - 2 = 0$ e $\frac{\partial \mathcal{L}}{\partial \lambda} = 3 \cdot x + 2 \cdot y + 5 \cdot z - 1 = 0$. As soluções das equações acima resultam no ponto $P^*\left(\frac{1}{4}, \frac{3}{4}, -\frac{1}{4}\right)$ e o raio da esfera é $r^* = \frac{\sqrt{35}}{4}$, que é a menor distância entre o plano e o ponto $C(1, 1, 1)$. Observe o fato óbvio que o ponto sobre o plano mais próximo de $C(1, 1, 1)$ é a **projeção ortogonal** do $C(1, 1, 1)$ sobre o plano, ou seja, o extremo $P^*\left(\frac{1}{4}, \frac{3}{4}, -\frac{1}{4}\right)$ sobre o plano do segmento ortogonal ao plano unindo $C(1, 1, 1)$ e P^* .

Exemplo 18:

$$\begin{array}{ll} \min/\max & x^2 + y^2 + z^2 \\ \text{sa} & x^2 + y^2 - z = 3 \\ & x + y + z = 2 \end{array}$$

Observe que $x^2 + y^2 + z^2$ é a distância ao quadrado de um ponto com coordenadas $P(x, y, z)$ à origem $(0, 0, 0)$. A restrição $x^2 + y^2 - z = 3$, $x + y + z = 2$ é a elipse na interseção do paraboloide $x^2 + y^2 - z = 3$ com o plano $x + y + z = 2$, conforme o desenho abaixo:



Você não precisa fazer os cálculos abaixo, mas acompanhe como o problema está sendo desenvolvido. O lagrangeano desse problema é:

$$\mathcal{L} = x^2 + y^2 + z^2 - \lambda_1 \cdot (x^2 + y^2 - z - 3) - \lambda_2 \cdot (x + y + z - 2).$$

As condições de 1ª ordem são

$\frac{\partial \mathcal{L}}{\partial x} = 2 \cdot x - 2 \cdot \lambda_1 \cdot x - \lambda_2 = 0$, $\frac{\partial \mathcal{L}}{\partial y} = 2 \cdot y - 2 \cdot \lambda_1 \cdot y - \lambda_2 = 0$, $\frac{\partial \mathcal{L}}{\partial z} = 2 \cdot z + \lambda_1 - \lambda_2 = 0$, disso obtemos $x = \frac{\lambda_2}{2 \cdot (1 - \lambda_1)}$, $y = \frac{\lambda_2}{2 \cdot (1 - \lambda_1)}$ e $z = \frac{\lambda_2 - \lambda_1}{2}$. Utilizando as restrições, obtemos:

$$\left(\frac{\lambda_2}{\lambda_1 - 1} \right)^2 + (\lambda_1 - \lambda_2) = 6$$

$$2 \cdot \left(\frac{\lambda_2}{\lambda_1 - 1} \right) + (\lambda_1 - \lambda_2) = -4.$$

Substituindo $a = \left(\frac{\lambda_2}{\lambda_1 - 1} \right)$ e $b = (\lambda_1 - \lambda_2)$, e resolvendo as equações

$$a^2 + b = 6$$

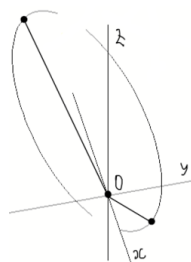
$$2 \cdot a + b = -4.$$

Obtemos dois valores: $a = \frac{\lambda_2}{\lambda_1 - 1} = 1 - \sqrt{11}$ e $a = \frac{\lambda_2}{\lambda_1 - 1} = 1 + \sqrt{11}$. Se $\frac{\lambda_2}{\lambda_1 - 1} = 1 - \sqrt{11}$, então $(1 - \sqrt{11}) \cdot \lambda_1 - \lambda_2 = 1 - \sqrt{11}$ e $\lambda_1 - \lambda_2 = -6 + 2 \cdot \sqrt{11}$, obtemos $\lambda_1 = 3 - \frac{7 \cdot \sqrt{11}}{11}$ e $\lambda_2 = 9 - \frac{29 \cdot \sqrt{11}}{11}$. Se $\frac{\lambda_2}{\lambda_1 - 1} = 1 + \sqrt{11}$, obtemos $\lambda_1 = 3 + \frac{7 \cdot \sqrt{11}}{11}$ e $\lambda_2 = 9 + \frac{29 \cdot \sqrt{11}}{11}$. Como $x = \frac{\lambda_2}{2 \cdot (1 - \lambda_1)}$, $y = \frac{\lambda_2}{2 \cdot (1 - \lambda_1)}$ e $z = \frac{\lambda_2 - \lambda_1}{2}$, utilizando os valores anteriores dos multiplicadores, obtemos dois pontos:

$$x = \frac{\sqrt{11} - 1}{2}, y = \frac{\sqrt{11} - 1}{2} \text{ e } z = 3 - \sqrt{11};$$

$$x = -\frac{\sqrt{11} + 1}{2}, y = -\frac{\sqrt{11} + 1}{2} \text{ e } z = 3 + \sqrt{11}.$$

O primeiro ponto corresponde ao ponto mais próximo da origem. Veja o desenho abaixo:



O lagrangeano determina esferas centradas na origem que tangenciam a elipse:

$$\begin{aligned}x^2 + y^2 - z &= 3 \\x + y + z &= 2,\end{aligned}$$

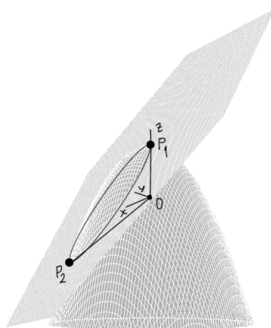
No ponto mais próximo, o raio da circunferência é $r = \sqrt{26 - 7\sqrt{11}}$, e para o ponto mais distante o raio é $r = \sqrt{26 + 7\sqrt{11}}$. Portanto, o valor mínimo para o problema proposto é $r^2 = 26 - 7\sqrt{11}$ e o valor máximo é $r^2 = 26 + 7\sqrt{11}$.

Exercício 5

Encontre os valores mínimo e máximo do problema:

$$\begin{aligned}\min/\max \quad & x^2 + y^2 + z^2 \\ \text{sa} \quad & x^2 + y^2 + 2 \cdot z = 2 \\ & x + y + z = 1.\end{aligned}$$

Primeiramente, observe que a restrição é uma elipse sobre a qual estão os pontos $P_1(0,0,1)$ e $P_2(2,2,-3)$ que distam 1 e $\sqrt{17}$ da origem $(0,0,0)$, respectivamente.

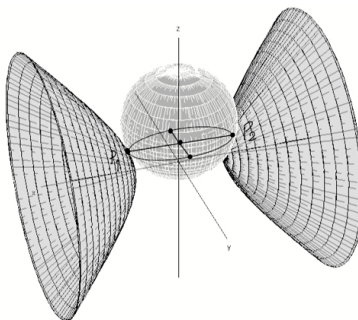


Formule o lagrangeano desse problema e verifique que $P_1(0,0,1)$, $\lambda_1=1$, $\lambda_2=0$ e $P_2(2,2,-3)$, $\lambda_1=5$, $\lambda_2=-16$ satisfazem as condições de primeira ordem do lagrangeano. $P_1(0,0,1)$ e $P_2(2,2,-3)$ correspondem ao ponto mais próximo à origem e ao ponto mais distante à origem sobre a elipse. As esferas centradas nesses pontos, com raios iguais a 1 e $\sqrt{17}$, tangenciam a elipse. Verifique que, nos pontos $P_1(0,0,1)$ e $P_2(2,2,-3)$, o gradiente de $f = x^2 + y^2 + z^2$ é combinação linear dos gradientes de $g_1 = x^2 + y^2 + 2 \cdot z = 2$ e $g_2 = x + y + z = 1$.

Exercício 6

Encontre o raio da esfera $x^2 + y^2 + (z-1)^2 = r^2$ centrada no ponto $P(0,0,1)$ que tangencia o hiperboloide $x^2 - y^2 - z^2 - 1 = 0$. Observe que a solução desse problema corresponde à solução do problema:

$$\begin{aligned} \min \quad & x^2 + y^2 + (z-1)^2 \\ \text{sa} \quad & x^2 - y^2 - z^2 - 1 = 0. \end{aligned}$$

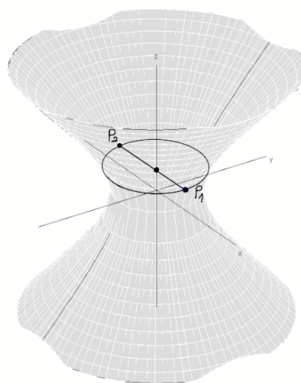


As soluções desses problemas estão nos pontos $P_1\left(\frac{\sqrt{5}}{2}, 0, \frac{1}{2}\right)$ e $P_2\left(-\frac{\sqrt{5}}{2}, 0, \frac{1}{2}\right)$, que estão sobre a esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{2}$ que tangencia o hiperboloide nesses pontos.

Exercício 7

Encontre o raio da esfera $x^2 + y^2 + (z-1)^2 = r^2$ centrada no ponto $P(0,0,1)$ que tangencia o hiperboloide $x^2 + y^2 - z^2 - 1 = 0$. Observe que a solução desse problema corresponde à solução do problema:

$$\begin{aligned} \min \quad & x^2 + y^2 + (z-1)^2 \\ \text{sa} \quad & x^2 + y^2 - z^2 - 1 = 0. \end{aligned}$$



As soluções desse problema estão sobre a circunferência $x^2 + y^2 = \frac{6}{4}$ e $z = \frac{1}{2}$. Os pontos $P_1\left(\frac{\sqrt{5}}{2}, 0, \frac{1}{2}\right)$ e $P_2\left(-\frac{\sqrt{5}}{2}, 0, \frac{1}{2}\right)$ estão sobre a circunferência. A esfera $x^2 + y^2 + (z-1)^2 = \frac{3}{2}$ tangencia o hiperboloide ao longo da circunferência $x^2 + y^2 = \frac{6}{4}$ e $z = \frac{1}{2}$.

Exercício 8

Encontre o ponto $P^*(x^*, y^*, z^*)$, sobre a reta definida pelos planos

$$g_1(x, y, z) = 2 \cdot x + y + z = 1 \text{ e } g_2(x, y, z) = x + 2 \cdot y + z = 2$$

que está mais próximo da origem $(0, 0, 0)$.

Exercício 9

Como as condições de 1ª ordem do lagrangeano expressam uma condição geométrica – a tangência – temos que utilizar argumento adicional para determinar se um ponto de tangência é um mínimo ou máximo local. Portanto, nos exercícios abaixo, você deve tomar consciência desses argumentos adicionais para responder o que está sendo solicitado:

i) Determine o ponto de mínimo e máximo global de $f(x, y) = 8 \cdot x^2 + 4 \cdot y + 4 \cdot y^2$ sujeito à restrição $x^2 + y^2 = 1$. Resposta: $(0, -1)$ é mínimo global e $\left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$ e $\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$ são máximos globais.

ii) Determine o ponto de mínimo e máximo global de $x+2 \cdot y+z$ sujeito à restrição $x^2+4 \cdot y^2+z^2=1$. Resposta: $\left(-\frac{1}{\sqrt{3}}, -\frac{1}{2\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$ é mínimo global e $\left(\frac{1}{\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ é máximo global.

iii) Encontrar a distância que o ponto $(0,0,1)$ está do paraboloide elíptico $x^2+y^2=z$. Resposta: $x^2+y^2=\frac{1}{2}$, $z=\frac{1}{2}$. Observe que $(0,0,0)$ é um ponto de máximo local para $x^2+y^2+(z-1)^2$ s.a. $x^2+y^2=z$.

iv) Determine o ponto mais próximo e o ponto mais distante sobre a esfera $x^2+y^2+z^2=36$ e o plano $x+y+z=6$ em relação ao ponto $(0,0,1)$ Resposta: $(0,0,6)$, $(4,4,-2)$.

Seja a superfície s definida por:

$$\begin{aligned} g_1(x_1, x_2, \dots, x_n) &= c_1, \\ g_2(x_1, x_2, \dots, x_n) &= c_2, \\ &\vdots \\ g_k(x_1, x_2, \dots, x_n) &= c_k. \end{aligned}$$

Suponhamos que, definida sobre a superfície s acima, tenhamos uma função $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ e que $x^* = (x_1^*, x_2^*, \dots, x_n^*) \in S$ seja um ponto de mínimo ou máximo local de f sobre s . Se $\tilde{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ é uma curva sobre s , tal que $\tilde{x}(0) = (x_1(0), x_2(0), \dots, x_n(0)) = x^*$, então

$$\frac{d}{dt} f(\tilde{x}(t))_{t=0} = \frac{d}{dt} f(x_1(t), x_2(t), \dots, x_n(t))_{t=0} = \nabla f(\tilde{x}(0)) \cdot \tilde{x}'(0) = 0,$$

isto é, $\nabla f(x^*) \cdot \tilde{x}'(0) = 0$, portanto, $\nabla f(x^*)$ é ortogonal a todo vetor tangente à superfície s em x^* . **Isso significa que $\nabla f(x^*)$ está no espaço ortogonal gerado pelos vetores $\nabla g_1(x^*), \dots, \nabla g_k(x^*)$** ¹, portanto, existem $\lambda_1^*, \dots, \lambda_k^* \in \mathbb{R}$ tais que

$$\nabla f(x^*) = \lambda_1^* \cdot \nabla g_1(x^*) + \lambda_2^* \cdot \nabla g_2(x^*) + \dots + \lambda_k^* \cdot \nabla g_k(x^*).$$

Em outras palavras,

$$\frac{\partial f}{\partial x_i}(x^*) = \lambda_1^* \cdot \frac{\partial g_1}{\partial x_i}(x^*) + \lambda_2^* \cdot \frac{\partial g_2}{\partial x_i}(x^*) + \dots + \lambda_k^* \cdot \frac{\partial g_k}{\partial x_i}(x^*), \text{ para } i=1, \dots, n.$$

Se definimos o lagrangeano como

$$\mathcal{L}(x, \lambda) = L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k) =$$

$$f(x_1, \dots, x_n) - \lambda_1 \cdot (g_1(x_1, \dots, x_n) - c_1) - \dots - \lambda_k \cdot (g_k(x_1, \dots, x_n) - c_k) =$$

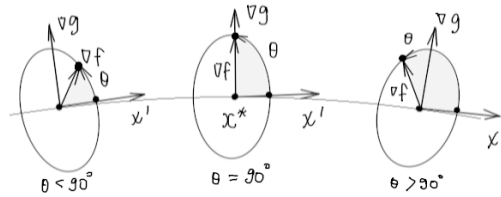
$$f(x) - \lambda_1 \cdot (g_1(x) - c_1) - \dots - \lambda_k \cdot (g_k(x) - c_k),$$

as condições acima se expressam como

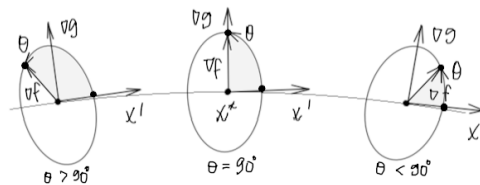
$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1}(x^*, \lambda^*) &= \frac{\partial f}{\partial x_1}(x^*) - \lambda_1^* \cdot \frac{\partial g_1}{\partial x_1}(x^*) - \lambda_2^* \cdot \frac{\partial g_2}{\partial x_1}(x^*) - \dots - \lambda_k^* \cdot \frac{\partial g_k}{\partial x_1}(x^*) = 0; \\ &\vdots \\ \frac{\partial \mathcal{L}}{\partial x_n}(x^*, \lambda^*) &= \frac{\partial f}{\partial x_n}(x^*) - \lambda_1^* \cdot \frac{\partial g_1}{\partial x_n}(x^*) - \lambda_2^* \cdot \frac{\partial g_2}{\partial x_n}(x^*) - \dots - \lambda_k^* \cdot \frac{\partial g_k}{\partial x_n}(x^*) = 0; \\ \frac{\partial \mathcal{L}}{\partial \lambda_1}(x^*, \lambda^*) &= -(g_1(x^*) - c_1) = 0; \\ &\vdots \\ \frac{\partial \mathcal{L}}{\partial \lambda_k}(x^*, \lambda^*) &= -(g_k(x^*) - c_k) = 0. \end{aligned}$$

Observe que a condição acima está afirmando que se x^* é um ponto de mínimo ou máximo local de f sobre s , então x^* pertence a uma isoquanta de f que tangencia a restrição em x^* , pois todos os vetores tangentes à s em x^* estão no plano tangente à isoquanta.

No desenho abaixo, a curva s está sobre a superfície $g(x_1, \dots, x_n) = c$ de dimensão $n-1$, isto é, $g(x(t)) = c$. Suponhamos que x^* é um máximo local para uma função f definida sobre a superfície $g(x_1, \dots, x_n) = c$. Na proximidade de x^* , a derivada $\frac{d}{dt}f(x(t))$, portanto, o ângulo θ , à medida que nos aproximamos de x^* , essa derivada diminui em valor absoluto se aproximando de zero. No ponto de máximo x^* , $\frac{d}{dt}f(x(0)) = \nabla f(x^*) \cdot x'(0) = |\nabla f(x^*)| \cdot |x'(0)| \cdot \cos \theta = 0$ e $\theta = 90^\circ$. Sendo assim, em x^* , os dois gradientes são paralelos, isto é, $\nabla f(x^*) = \lambda^* \cdot \nabla g(x^*)$, pois ambos são perpendiculares a todos os vetores tangentes à superfície $g(x_1, \dots, x_n) = c$ e esses vetores estão sobre um espaço afim de dimensão $n-1$. Conforme nos afastamos de x^* , o valor de $\frac{d}{dt}f(x(t))$ diminui, desse modo, $\frac{d}{dt}f(x(t)) = \nabla f(x(t)) \cdot x'(t) = |\nabla f(x(t))| \cdot |x'(t)| \cdot \cos \theta < 0$ e $\theta < 90^\circ$.

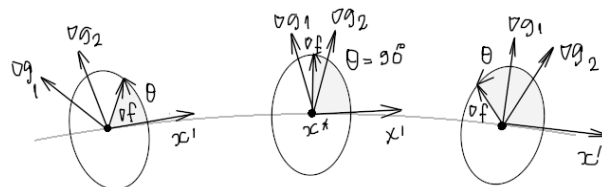


Caso x^* seja um mínimo local, quando nos aproximamos de x^* , a derivada $\frac{d}{dt}f(x(t)) = \nabla f(x(t)) \cdot x'(t) = |\nabla f(x(t))| \cdot |x'(t)| \cdot \cos \theta < 0$, portanto o ângulo θ , à medida que nos aproximamos de x^* , essa derivada diminui em valor absoluto se aproximando de zero. No ponto de mínimo x^* , $\frac{d}{dt}f(x(0)) = \nabla f(x^*) \cdot x'(0) = |\nabla f(x^*)| \cdot |x'(0)| \cdot \cos \theta = 0$ e $\theta = 90^\circ$. Dessa forma, em x^* , os dois gradientes são paralelos, isto é, $\nabla f(x^*) = \lambda \cdot \nabla g(x^*)$, pois ambos são perpendiculares a todos os vetores tangentes à superfície $g(x_1, \dots, x_n) = c$. Conforme nos afastamos de x^* , o valor de θ aumenta, por conseguinte, $\cos \theta$ diminui. Veja o desenho abaixo:



Suponhamos que a função f esteja definida sobre a superfície $S = g_1 \cap g_2$ obtida pela interseção de duas superfícies, g_1 e g_2 , cada uma delas de dimensão $n-1$. Seja a curva γ sobre as duas superfícies, isto é, $\gamma(t) \in S$ e $g_2(\gamma(t)) = c_2$. Portanto, os vetores $\gamma'(t)$ estão no espaço afim de dimensão $n-1$, que é tangente à superfície S em cada ponto $\gamma(t)$. Os gradientes $\nabla g_1(\gamma(t))$ e $\nabla g_2(\gamma(t))$ são ortogonais aos $\gamma'(t)$ e estão no espaço ortogonal ao plano tangente, que tem dimensão dois. Caso x^* seja um máximo local para f em S , quando nos aproximamos de x^* , a derivada $\frac{d}{dt}f(\gamma(t))$, portanto o ângulo θ , à medida que nos aproximamos de x^* , essa derivada diminui em valor absoluto, se aproximando de zero. No ponto de máximo x^* , $\frac{d}{dt}f(\gamma(0)) = \nabla f(x^*) \cdot \gamma'(0) = |\nabla f(x^*)| \cdot |\gamma'(0)| \cdot \cos \theta = 0$ e $\theta = 90^\circ$. Dessa forma, em x^* , $\nabla f(x^*)$ está no plano ortogonal ao espaço tangente em x^* , como $\nabla g_1(x)$ e $\nabla g_2(x)$ são linearmente

independentes então $\nabla f(x^*) = \lambda_1^* \cdot \nabla g_1(x^*) + \lambda_2^* \cdot \nabla g_2(x^*)$. Conforme nos afastamos de x^* , o valor de θ diminui, consequentemente, $\frac{d}{dt} f(x(t)) = \nabla f(x(t)) \cdot x'(t) = |\nabla f(x(t))| \cdot |x'(t)| \cdot \cos \theta < 0$ e $\theta < 90^\circ$. Veja o desenho abaixo:



Caso x^* seja um mínimo local para f em $g_1 \cap g_2$, quando nos aproximamos de x^* , a derivada $\frac{d}{dt} f(x(t)) = \nabla f(x(t)) \cdot x'(t) = |\nabla f(x(t))| \cdot |x'(t)| \cdot \cos \theta < 0$, portanto o ângulo θ , à medida que nos aproximamos de x^* , essa derivada diminui em valor absoluto se aproximando de zero. No ponto de máximo x^* , $\frac{d}{dt} f(x(0)) = \nabla f(x^*) \cdot x'(0) = |\nabla f(x^*)| \cdot |x'(0)| \cdot \cos \theta = 0$ e $\theta = 90^\circ$. Desse modo, em x^* , $\nabla f(x^*)$ está no plano ortogonal ao espaço tangente em x^* , como $\nabla g_1(x)$ e $\nabla g_2(x)$ são linearmente independentes, então $\nabla f(x^*) = \lambda_1^* \cdot \nabla g_1(x^*) + \lambda_2^* \cdot \nabla g_2(x^*)$. Conforme nos afastamos de x^* , o valor de θ aumenta, portanto, $\frac{d}{dt} f(x(t)) > 0$. Veja o desenho abaixo:

¹ Neste ponto, é essencial estar ciente que o subespaço dos vetores tangentes é o ortogonal do conjunto dos gradientes $\nabla g_1(x^*), \dots, \nabla g_k(x^*)$.

Lição CINCO

O Teorema de Kuhn-Tucker

Sumário

Lição 5 – O Teorema de Kuhn-Tucker

1. Introdução
2. Trajetórias admissíveis
3. O Teorema de Kuhn-Tucker
4. Lema de Farkas
5. A demonstração I do Teorema de Kuhn-Tucker
6. Demonstração II do Teorema de Kuhn-Tucker
 - 6.1 Uma versão do lagrangeano
7. O Teorema da Dualidade da Programação Linear
 - 7.1 Alternativas de Gale
 - 7.2 A Programação Linear, os lagrangeanos e as selas
8. O Teorema de Kuhn-Tucker vetorial
 - 8.1 O Teorema de Kuhn-Tucker para a otimização vetorial
9. Tucker e Kuhn

Lição 5

O Teorema de Kuhn-Tucker

1. Introdução

Nesta lição, nosso objetivo é apresentar o chamado Teorema de Kuhn-Tucker. Infere-se este resultado do celebrado artigo “*Nonlinear Programming*”¹, de 1950, publicado por Harold W. Kuhn (1925-2014) e Albert W. Tucker (1905-1995). Esse teorema é o tema central do nosso curso. Tudo que obtivemos anteriormente em programação linear pode ser adequadamente tratado com esse resultado. Tudo o que faremos nas próximas lições objetiva tornar as condições suficientes para mínimos e máximos condições necessárias expressas no Teorema de Kuhn-Tucker.

Ao estabelecerem o resultado principal do artigo, Kuhn e Tucker enunciaram as condições necessárias para um mínimo ou máximo local para otimização restrita, conhecidas como **condições de Kuhn-Tucker** (K-T), que serão tema da presente lição. Essas condições devem ser satisfeitas por pontos de mínimos e máximos locais de uma função objetivo $f(x)$, $x \in \mathbb{R}^n$, sujeita a um conjunto Ω de restrições definido por desigualdades $g_i(x) \leq a_i$ e $h_j(x) \geq b_j$:

$$\begin{array}{ll} \min / \max & f(x_1, \dots, x_n) \\ \text{sa} & g_1(x_1, \dots, x_n) \leq a_1; \\ & \vdots \\ & g_k(x_1, \dots, x_n) \leq a_k; \\ & h_1(x_1, \dots, x_n) \geq b_1; \\ & \vdots \\ & h_l(x_1, \dots, x_n) \geq b_l. \end{array}$$

Não era o objetivo dos autores obter esse resultado, que já havia sido estabelecido por outros autores, como o próprio Kuhn² salienta. De fato, no artigo “*Nonlinear Programming*”, o objetivo dos autores é investigar como a dualidade nos programas lineares se transfigura na programação não-linear. Kuhn e Tucker consideraram o problema:

$$\begin{array}{ll} \max & g(x_1, \dots, x_n) \\ \text{sa} & f_1(x_1, \dots, x_n) \geq 0; \\ & \vdots \\ & f_m(x_1, \dots, x_n) \geq 0; \\ & x_1 \geq 0, \dots, x_n \geq 0, \end{array}$$

e mostraram que (x^*, u^*) – em que x^* é um máximo global e u^* é o vetor formado com os respectivos multiplicadores de Lagrange – é um ponto de sela do lagrangeano $\phi(x, u) = g(x) + u_1 \cdot f_1(x) + \dots + u_m \cdot f_m(x)$, ou seja, $\phi(x, u^*) \leq \phi(x^*, u^*) \leq \phi(x^*, u)$, desde que as funções $g, f_1, \dots, f_m$ sejam côncavas. Esse resultado se encontra no final da Lição 6.

Além da originalidade do resultado, outra novidade do artigo está na demonstração do teorema principal, empregando o chamado “**Lema de Farkas**”, um resultado do começo do século 20, redescoberto, aparentemente, por von Neumann no contexto do desenvolvimento da programação linear na década de 1940.

Nesta lição, seguimos a maneira aclamada: “Teorema de Kuhn-Tucker”, que é o resultado que estabelece as condições de Kuhn-Tucker para mínimos e máximos. Em relação a essas condições, outro autor já havia empregado o Lema de Farkas no contexto da otimização. O matemático William Karush (1917-1997) – antes de von Neumann, Kuhn e Tucker – já empregara o Lema na sua dissertação de mestrado, defendida na Universidade de Chicago em dezembro de 1939 e que não foi publicada. Resultado semelhante às condições de Kuhn e Tucker também está num artigo publicado, em 1948, pelo matemático Fritz John (1910-1994). O artigo é anterior a 1948 e foi inicialmente rejeitado para publicação³. Tendo em vista os antecedentes do teorema, alguns autores nomeiam o resultado

que nos interessa como “Kuhn-Tucker-Karush-Fritz John Theorem”. Um nome demasiadamente longo ainda que, do ponto de vista histórico, justo. Num artigo acerca do histórico do teorema, Kuhn analisa o contexto matemático dos resultados de Karush e Fritz John, que é bem diverso do contexto no qual Kuhn e Tucker investigavam as potencialidades da dualidade numa versão generalizada.⁴

O artigo de Kuhn-Tucker ainda generaliza o resultado obtido para a atualmente chamada “**otimização vetorial**”. Como vimos, a programação vetorial diferencia-se dos problemas acima por duas razões. A primeira delas é que temos várias funções-objetivo $f_1(x)$, ..., $f_p(x)$. A segunda, é que se faz necessário definir o que, nesse contexto, se entende por minimizar ou maximizar um conjunto de funções $f_1(x)$, ..., $f_p(x)$. A definição empregada é aquela conhecida dos economistas como um “Ótimo de Pareto”.

Como não poderia deixar de ser, Kuhn e Tucker também aplicaram os resultados obtidos à Programação Linear. Interpretaremos a Programação Linear como um caso particular do Teorema de Kuhn-Tucker. No caso da Programação Linear, e com o emprego do Teorema de Kuhn-Tucker, não é difícil mostrar que pares de soluções dos problemas primal e dual são selas do respectivo lagrangeano. Também mostramos que dado um dos problemas, quando resolvido pelo Método Simplex, a solução do outro problema, associada à solução obtida do primeiro, é dada pelos coeficientes das variáveis de folga, como vimos na Lição 3.

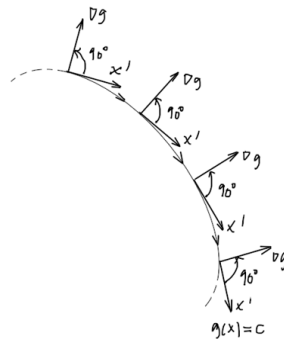
Os comentários acima resumem os fatos principais desse artigo que está entre os mais importantes da Matemática Aplicada do século 20.

2. Trajetórias admissíveis

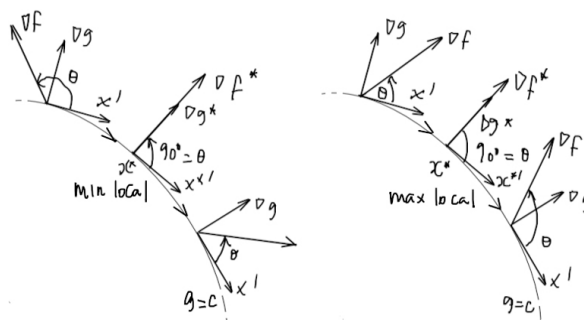
Já estudamos trajetórias anteriormente. Uma trajetória, ou curva, no $\mathbb{R}^n$, é uma função $\vec{x}: (t_0, t_1) \rightarrow \mathbb{R}^n$, $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_n(t)) \in \mathbb{R}^n$, tal que em cada ponto $t \in (t_0, t_1)$ existe $\vec{x}'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))$, dito vetor tangente à curva ou vetor velocidade em t .

Utilizamos tais trajetórias para obter as condições de 1ª ordem do lagrangeano.

Nas situações anteriores, evitamos tratar o fato que a trajetória $x(t)$ “atravessa” as isoquantas de $f(x)$ e enfatizamos o fato que estavam limitadas à isoquanta $g(x)=c$. Por exemplo, se determinada trajetória está contida na superfície $g(x)=c$, então $\frac{d}{dt}g(x(t)) = \nabla g(x(t)) \cdot x'(t) = |\nabla g(x(t))| \cdot |x'(t)| \cdot \cos(\theta(t)) = 0$, ou seja, o gradiente $\nabla g(x(t))$ e $x'(t)$ são ortogonais, como vemos no desenho abaixo:



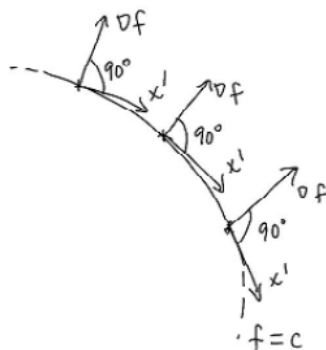
Caso tenhamos um ponto de mínimo ou máximo x^* de uma função $f(x)$, definida sobre $g(x)=c$, então a derivada $\frac{d}{dt}f(x(t)) = \nabla f(x(t)) \cdot x'(t)$ explica os ângulos entre os vetores $\nabla f(x(t))$ e $x'(t)$ numa vizinhança de x^* e por que $\nabla f(x^*)$ e $\nabla g(x^*)$ são ortogonais. Veja os desenhos abaixo:



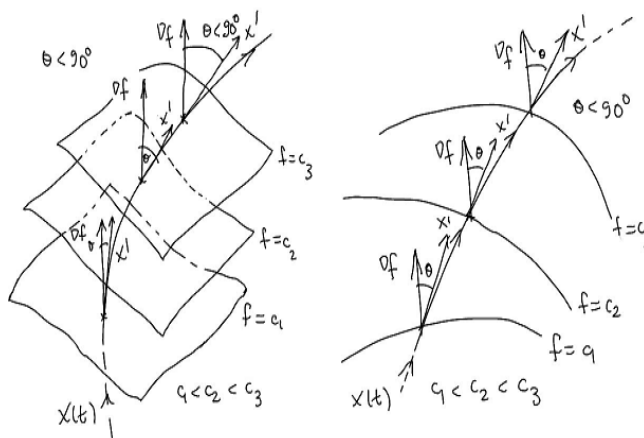
Esta é a primeira vez que empregamos a noção de “vizinhança”. Por vizinhança de um ponto $x^* \in \mathbb{R}^n$, entendemos as bolas abertas ou fechadas centradas em x^* , ou seja, os subconjuntos $B(x^*, \varepsilon) = \{x \in \mathbb{R}^n \mid \|x^* - x\| < \varepsilon\}$ e $\bar{B}(x^*, \varepsilon) = \{x \in \mathbb{R}^n \mid \|x^* - x\| \leq \varepsilon\}$, respectivamente. $B(x^*, \varepsilon)$ é constituído pelos pontos $x \in \mathbb{R}^n$ cuja distância entre x e x^* é menor que $\varepsilon > 0$, isto explica o nome “bola aberta”, $\varepsilon > 0$ é o raio da bola.

Por sua vez, $\bar{B}(x^*, \varepsilon)$ é constituído pelos pontos $x \in \mathbb{R}^n$ cuja distância entre x e x^* é menor ou igual a $\varepsilon \geq 0$. As bolas fechadas são as uniões das bolas abertas $B(x^*, \varepsilon)$ com a esfera $S(x^*, \varepsilon) = \{x \in \mathbb{R}^n \mid \|x^* - x\| = \varepsilon\}$. A bola $\bar{B}(x^*, 0)$ de raio $\varepsilon = 0$, se reduz ao ponto x^* . Nos referimos as bolas $B(x^*, \varepsilon)$ e $\bar{B}(x^*, \varepsilon)$ como vizinhanças abertas e fechadas, respectivamente, em torno de x^* . Em geral, quando é indiferente tomarmos uma bola aberta ou fechada, estaremos simplesmente nomeando “vizinhança”.

Consideremos, agora, a situação em que a trajetória $x(t)$ “atravessa” várias isoquantas de f . Repetindo: se a trajetória permanece na mesma isoquanta, o ângulo entre $\nabla f(x(t))$ e $x'(t)$ se conserva igual a 90° . Veja o desenho:



Caso o ângulo $\theta(t)$ permaneça menor que 90° , $\nabla f(x(t)) \cdot x'(t) = \|\nabla f(x(t))\| \cdot \|x'(t)\| \cdot \cos(\theta(t)) > 0$, estamos passando de isoquantas de nível menor $f=c_1$ para isoquantas de nível maior $f=c_3$. Veja os desenhos abaixo:



Em cada ponto $x(t)$ em que a velocidade é $|x'(t)|$, o maior valor possível para $\frac{d}{dt}f(x(t)) = \nabla f(x(t)) \cdot x'(t)$ está na direção e no sentido do gradiente $\nabla f(x(t))$, pois nessa direção e nesse sentido $\theta = 0$ e $\cos(\theta(t)) = 1$. Se o deslocamento se dá no sentido contrário ao gradiente, ou seja, $\theta = \pi$, então o valor de $\frac{d}{dt}f(x(t))$ é o menor possível, dado $x(t)$ e $|x'(t)|$. Observe que o valor $\frac{d}{dt}f(x(t)) = \nabla f(x(t)) \cdot x'(t) = |\nabla f(x(t))| \cdot |x'(t)| \cdot \cos(\theta(t))$ depende de $|\nabla f(x(t))|$, $|x'(t)|$ e $\theta(t)$.

Obviamente, caso estejamos simultaneamente “atravessando” isoquantas de f e g , as observações acima valem tanto para f quanto para g . E exatamente isso é o que faremos mais adiante.

Consideremos a região $\Omega \subseteq \mathbb{R}^n$ definida por

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l.$$

Um ponto $x \in \Omega$ é dito um **ponto viável**. Seja $x^* \in \mathbb{R}^n$ um ponto pertencente à região $\Omega \subseteq \mathbb{R}^n$. Dizemos que a fronteira $g_i(x) = a_i$ é **ativa** em x^* se $g_i(x^*) = a_i$; se $g_i(x^*) < a_i$, dizemos que a fronteira $g_i(x) = a_i$ é **inativa** em x^* . Da mesma maneira, a fronteira $h_j(x) = b_j$ é **ativa** em x^* se $h_j(x^*) = b_j$ etc.

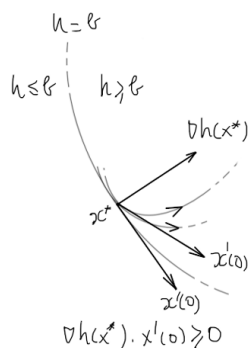
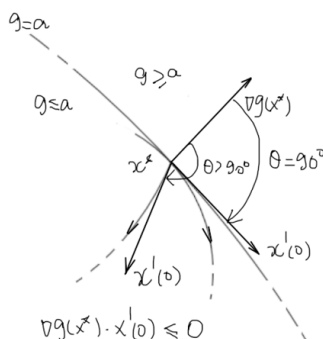
Uma **trajetória admissível em x^*** é uma curva derivável $x: [0, +\infty) \rightarrow \Omega$, tal que $x(0) = x^*$ e, caso as fronteiras $g_i(x) = a_i$ e $h_j(x) = b_j$ sejam ativas em x^* ,

$$\left. \frac{d}{dt}g_i(x(t)) \right|_{t=0} = \nabla g_i(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla g_i(x^*) \cdot x'(0) \leq 0;$$

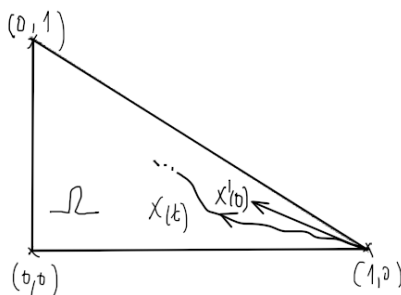
$$\left. \frac{d}{dt}h_j(x(t)) \right|_{t=0} = \nabla h_j(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla h_j(x^*) \cdot x'(0) \geq 0.$$

As condições acima estão nos assegurando que se a trajetória $x: [0, +\infty) \rightarrow \Omega$ inicia num ponto x^* sobre determinada fronteira, não o faz numa direção que “corta” essa fronteira saindo de Ω . Por exemplo, digamos que $g_i(x^*) = a_i$, então $\nabla g_i(x^*) \cdot x'(0) = |\nabla g_i(x^*)| \cdot |x'(0)| \cdot \cos \theta \leq 0$. Portanto, o ângulo θ entre $\nabla g_i(x^*)$ e $x'(0)$ é maior ou igual a 90° . Como $\nabla g_i(x^*)$ é perpendicular ao plano tangente à superfície $g_i(x) = a_i$ em x^* e aponta para dentro da região $g_i(x) \geq a_i$ e o ângulo entre $\nabla g_i(x^*)$ e $x'(0)$ é no mínimo 90° , então a curva $x(t)$ está tangenciando a isoquanta $g_i(x) = a_i$ ou entrando na

região $g_i(x) \leq a_i$. Da mesma maneira, deve-se interpretar $\nabla h_j(x^*) \cdot x'(0) \geq 0$. Observe os desenhos abaixo:



Quando a fronteira é inativa no ponto x^* , ou seja, $g_i(x^*) < a_i$ ou $h_j(x^*) > b_j$, não se faz necessária, em princípio, localmente, ou seja, na proximidade de x^* , nenhuma condição em relação às derivadas acima, pois nesse caso todas as direções $x'(0)$ são localmente admissíveis em relação a essa fronteira. Entretanto, isso é falso! Por exemplo, considere as restrições $h(x_1, x_2) = 1 - x_1 - x_2 \geq 0$, $h_1(x_1, x_2) = x_1 \geq 0$ e $h_2(x_1, x_2) = x_2 \geq 0$:



Necessariamente, para uma curva admissível, tal que $x(t) = (x_1(t), x_2(t))$ e $x(0) = (1, 0)$, $x'_1(0) \leq 0$, $x'_2(0) \geq 0$. Caso $x'_1(0) < 0$ e $x'_2(0) > 0$, como no desenho acima, para a curva não sair do triângulo, necessariamente $\frac{x'_2(0)}{|x'_1(0)|} \leq 1$, pois a declividade da reta $x_1 + x_2 = 1$ é -1 . Isso implica que $x'_1(0) + x'_2(0) \leq 0$ ou $x'_1(0) \leq -x'_2(0)$. Portanto, $\left. \frac{d}{dt} h_1(x_1(t), x_2(t)) \right|_{t=0} = x'_1(0) \leq -x'_2(0) < 0$. Observe que a fronteira $h_1(x_1, x_2) = x_1 \geq 0$ é inativa em $x(0) = (1, 0)$. Portanto, ainda que as fronteiras sejam inativas, condições sobre $\nabla g(x^*) \cdot x'(0)$ ou $\nabla h(x^*) \cdot x'(0)$ podem, todavia, existir, por imposição das demais restrições. Esse exemplo está no artigo de Kuhn e Tucker.

Nos problemas econômicos, normalmente temos as restrições $x_1 \geq 0, \dots, x_n \geq 0$. Também, em relação a essas fronteiras, as curvas admissíveis estão submetidas às condições acima. Caso essas restrições sejam representadas pelas projeções $\pi_1(x) = x_1 \geq 0, \dots, \pi_n(x) = x_n \geq 0$ e a fronteira $\pi_i(x) = x_i \geq 0$ seja ativa em x^* , então

$$\left. \frac{d}{dt} \pi_i(x(t)) \right|_{t=0} = \nabla \pi_i(x(t)) \cdot x'(t) \Big|_{t=0} = x'_i(0) \geq 0.$$

3. O Teorema de Kuhn-Tucker

Abaixo enunciamos o Teorema de Kuhn-Tucker. A hipótese chamada “**qualificação de restrição**” é esclarecida posteriormente. Por hora, basta sabermos que essa condição é satisfeita nos problemas que nos interessam.

Teorema 1 (Kuhn-Tucker 1950)

i) Se x^* é um máximo local para $f(x)$, sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_1 \geq 0, \dots, x_n \geq 0,$$

tal que x^* satisfaz a qualificação de restrição e $\nabla f(x^*) \neq 0$. Então, existem $\lambda_i^* \in \mathbb{R}, i=1, \dots, k$, $\mu_j^* \in \mathbb{R}, j=1, \dots, l$, $\delta_k^* \in \mathbb{R}, i=1, \dots, n$ que satisfazem as condições abaixo, ditas condições de Kuhn-Tucker para um máximo:

$$\begin{aligned} KT1: \quad & \nabla f(x^*) = \sum \lambda_i^* \cdot \nabla g_i(x^*) + \sum \mu_j^* \cdot \nabla h_j(x^*) + \sum \delta_k^* \cdot \vec{e}_k, \\ KT2: \quad & \lambda_i^* \cdot (g_i(x^*) - a_i) = 0, \quad \mu_j^* \cdot (h_j(x^*) - b_j) = 0, \quad \delta_k^* \cdot x_k^* = 0, \\ KT3: \quad & \lambda_i^* \geq 0, \quad \mu_j^* \leq 0, \quad \delta_k^* \leq 0. \end{aligned}$$

ii) Se x^* é um mínimo local para $f(x)$ sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, x_1 \geq 0, \dots, x_n \geq 0,$$

tal que x^* satisfaz a qualificação de restrição e $\nabla f(x^*) \neq 0$. Então, existem $\lambda_i^* \in \mathbb{R}, i=1, \dots, k, \mu_j^* \in \mathbb{R}, j=1, \dots, l, \nu_k^* \in \mathbb{R}, k=1, \dots, n$ que satisfazem as condições abaixo, ditas condições de Kuhn-Tucker para um mínimo:

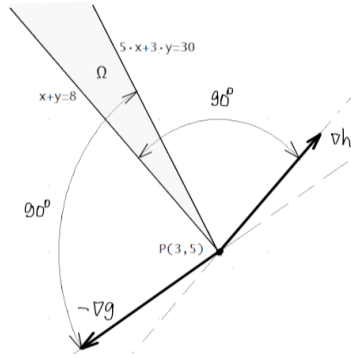
$$\begin{aligned} KT1: \quad \nabla f(x^*) &= \sum \lambda_i^* \cdot \nabla g_i(x^*) + \sum \mu_j^* \cdot \nabla h_j(x^*) + \sum \nu_k^* \cdot \vec{e}_k, \\ KT2: \quad \lambda_i^* \cdot (g_i(x^*) - a_i) &= 0, \quad \mu_j^* \cdot (h_j(x^*) - b_j) = 0, \quad \nu_k^* \cdot x_k^* = 0, \\ KT3: \quad \lambda_i^* &\leq 0, \quad \mu_j^* \geq 0, \quad \nu_k^* \geq 0. \end{aligned}$$

Heuristicamente, compreendamos o Teorema de Kuhn-Tucker. Vamos supor que apenas duas restrições são ativas, o que não implica perda de generalidade.

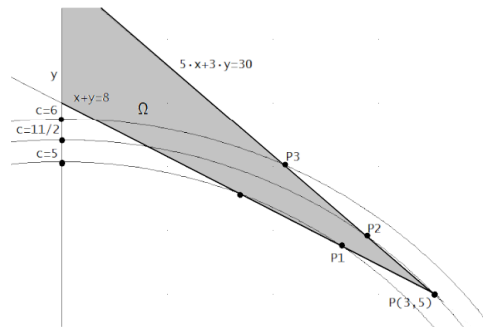
Exemplo 1: Consideremos o problema:

$$\begin{aligned} \max \quad & \frac{x^2}{3} + \frac{y^2}{10} \\ \text{sa} \quad & 5 \cdot x + 3 \cdot y \leq 30 \\ & x + y \geq 8. \end{aligned}$$

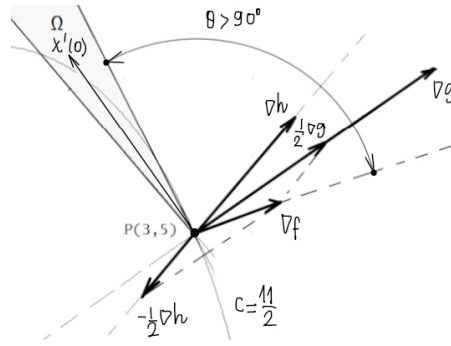
Sejam $f(x, y) = \frac{x^2}{3} + \frac{y^2}{10}$, $g(x, y) = 5 \cdot x + 3 \cdot y$ e $h(x, y) = x + y$. A região Ω delimitada pelas restrições $g(x, y) \leq 30$ e $h(x, y) \geq 8$ está desenhada abaixo e corresponde ao conjunto de semirretas, $t \geq 0$, com origem no ponto $P(3, 5)$ na direção dos vetores $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v \in \mathbb{R}^2$, tais que $-\nabla g(3, 5) \cdot v \geq 0$ e $\nabla h(3, 5) \cdot v \geq 0$, ou seja, $\Omega = \left\{ \begin{pmatrix} 3 \\ 5 \end{pmatrix} + t \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2 \mid t \geq 0, -\nabla g(3, 5) \cdot v \geq 0, \nabla h(3, 5) \cdot v \geq 0 \right\}$. Esse é um exemplo daquilo que definimos como um “cone”. O ponto $P(3, 5)$ é o vértice desse cone que, obviamente, está na interseção das fronteiras $g(x, y) = 30$ e $h(x, y) = 8$.



Observe que no ponto $P(3,5)$, $\nabla f(3,5) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 3 \end{pmatrix} + \begin{pmatrix} -1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \cdot \nabla g(3,5) + \mu \cdot \nabla h(3,5)$, ou seja, $P(3,5)$ satisfaz as condições de Kuhn-Tucker para um máximo local. Observe o desenho abaixo e convença-se que, de fato, é um máximo local. A isoquanta de $f(x,y) = \frac{x^2}{3} + \frac{y^2}{5}$ que passa por esse ponto é $\frac{x^2}{3} + \frac{y^2}{10} = \frac{11}{2}$. Existem isoquantas de $f(x,y) = \frac{x^2}{3} + \frac{y^2}{5}$, de nível maior que $\frac{11}{2}$, que interceptam Ω , por exemplo, $\frac{x^2}{3} + \frac{y^2}{10} = 6$, mas todas interceptam Ω em pontos sobre a reta $5 \cdot x + 3 \cdot y = 30$ que estão à esquerda de P . A isoquanta $\frac{x^2}{3} + \frac{y^2}{10} = 6$ intercepta $5 \cdot x + 3 \cdot y = 30$ no ponto P_3 .



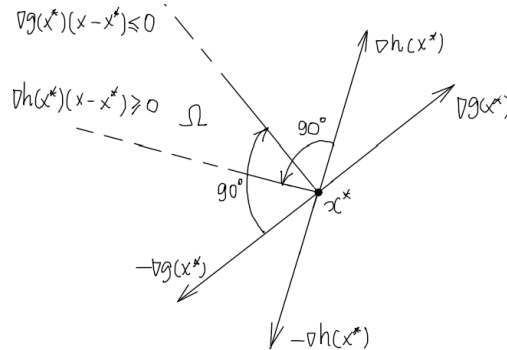
Observe que o ângulo entre o vetor $\nabla f(3,5)$ e qualquer vetor $v \in \mathbb{R}^2$, tal que $-\nabla g(3,5) \cdot v \geq 0$ e $\nabla h(3,5) \cdot v \geq 0$ é maior que 90° , portanto, se $x(t)$ é um caminho viável, $x(0) = (3,5)$, $x'(0) = v$, então $\nabla f(3,5) \cdot x'(0) < 0$ e essa última expressão está assegurando que $P(3,5)$ é um máximo local.



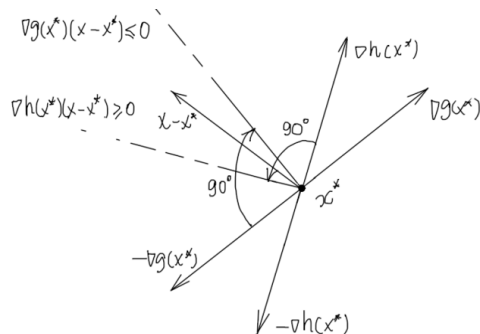
Exemplo 2: Consideremos a generalização no $\mathbb{R}^n$ do problema anterior:

$$\begin{aligned} \max \quad & f(x) \\ \text{sa} \quad & g(x^*) \cdot (x - x^*) \leq 0 \\ & h(x^*) \cdot (x - x^*) \geq 0 \end{aligned}$$

O cone $\Omega = \{x^* + t \cdot (x - x^*) \mid t \geq 0, -\nabla g(x^*) \cdot (x - x^*) \geq 0, \nabla h(x^*) \cdot (x - x^*) \geq 0\}$ está desenhado abaixo e corresponde ao conjunto das semirretas, $t \geq 0$, com origem no ponto x^* na direção dos vetores $(x - x^*)$, tais que $-\nabla g(x^*) \cdot (x - x^*) \geq 0$ e $\nabla h(x^*) \cdot (x - x^*) \geq 0$. O vértice x^* do cone está na interseção das fronteiras $\nabla g(x^*) \cdot (x - x^*) = 0$ e $\nabla h(x^*) \cdot (x - x^*) = 0$.



Se a trajetória $x(t)$ é admissível, então os ângulos θ_f , θ_g e θ_h que o vetor velocidade $x'(0) = x - x^*$ faz com os gradientes de f , g e h , respectivamente, são tais que $\cos \theta_f \leq 0$, $\cos \theta_g \leq 0$ e $\cos \theta_h \geq 0$. Veja o desenho abaixo. Como $\nabla f(x^*) \cdot x'(0) = \lambda^* \cdot \nabla g(x^*) \cdot x'(0) + \mu^* \cdot \nabla h(x^*) \cdot x'(0)$, então $\|\nabla f(x^*)\| \cdot \|x'(0)\| \cdot \cos \theta_f = \lambda^* \cdot \|\nabla g(x^*)\| \cdot \|x'(0)\| \cdot \cos \theta_g + \mu^* \cdot \|\nabla h(x^*)\| \cdot \|x'(0)\| \cdot \cos \theta_h$. Se $x'(0) \neq 0$, então $\|\nabla f(x^*)\| \cdot \cos \theta_f = \lambda^* \cdot \|\nabla g(x^*)\| \cdot \cos \theta_g + \mu^* \cdot \|\nabla h(x^*)\| \cdot \cos \theta_h$. Como podemos variar esses ângulos desde que $\cos \theta_f \leq 0$, $\cos \theta_g \leq 0$ e $\cos \theta_h \geq 0$, então só é possível manter $\|\nabla f(x^*)\| \cdot \cos \theta_f = \lambda^* \cdot \|\nabla g(x^*)\| \cdot \cos \theta_g + \mu^* \cdot \|\nabla h(x^*)\| \cdot \cos \theta_h$ se $\lambda^* \geq 0$ e $\mu^* \leq 0$. Por exemplo, se $\theta_g = 90^\circ$, $\|\nabla f(x^*)\| \cdot \cos \theta_f = \mu^* \cdot \|\nabla h(x^*)\| \cdot \cos \theta_h$, então $\mu^* \leq 0$; se $\theta_h = 90^\circ$, $\|\nabla f(x^*)\| \cdot \cos \theta_f = \lambda^* \cdot \|\nabla g(x^*)\| \cdot \cos \theta_g$, então $\lambda^* \geq 0$.



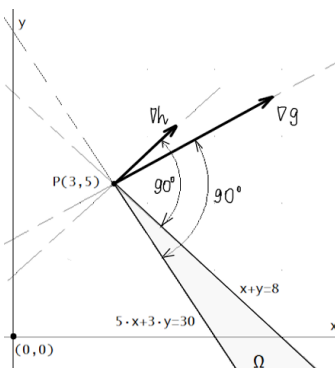
Portanto, mostramos que, se x^* é um máximo local, então $\lambda^* \geq 0$ e $\mu^* \leq 0$.

No desenho acima, supomos que a região Ω é a interseção dos semiespaços $\nabla g(x^*) \cdot (x - x^*) \leq 0$ e $\nabla h(x^*) \cdot (x - x^*) \geq 0$. Nesse caso, para qualquer $v \in \mathbb{R}^n$, tal que $-\nabla g(x^*) \cdot v \geq 0$ e $\nabla h(x^*) \cdot v \geq 0$, existe uma curva admissível $x'(0) = v$. Para a prova do Teorema de Kuhn-Tucker, os autores introduziram a hipótese que todo o vetor v no cone determinado por $-\nabla g(x^*) \cdot v \geq 0$ e $\nabla h(x^*) \cdot v \geq 0$ é tal que existe uma curva admissível em Ω , isto é, $x(0) = x^*$ e $x'(0) = v$. Essa hipótese é chamada “Qualificação de Restrição”⁵.

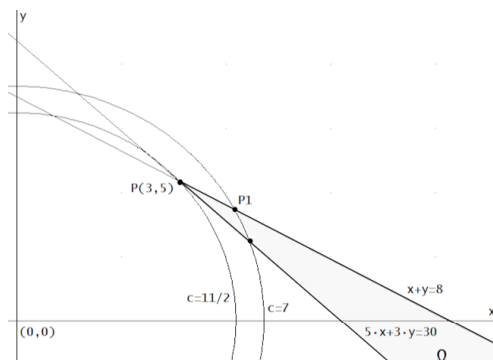
Exemplo 3: Consideremos o problema:

$$\begin{aligned} \min \quad & \frac{x^2}{3} + \frac{y^2}{10} \\ \text{sa} \quad & 5 \cdot x + 3 \cdot y \geq 30 \\ & x + y \leq 8. \end{aligned}$$

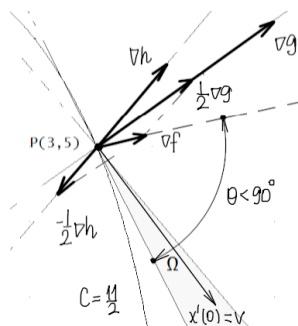
O cone que estamos considerando é a região Ω delimitada pelas restrições $g(x, y) \geq 30$ e $h(x, y) \leq 8$, está desenhado abaixo e corresponde ao conjunto de semirretas, $t \geq 0$, com origem no ponto $P(3, 5)$ na direção dos vetores $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v \in \mathbb{R}^2$, tais que $\nabla g(3, 5) \cdot v \geq 0$ e $\nabla h(3, 5) \cdot v \geq 0$, ou seja, $\Omega = \left\{ \begin{pmatrix} 3 \\ 5 \end{pmatrix} + t \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2 \mid t \geq 0, \nabla g(3, 5) \cdot v \geq 0, \nabla h(3, 5) \cdot v \geq 0 \right\}$.



Observe que $\nabla f(3,5) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 3 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \mu \cdot \nabla g(3,5) + \lambda \cdot \nabla h(3,5)$, ou seja, $P(3,5)$ satisfaz as condições de Kuhn-Tucker para um mínimo local. Observe o desenho abaixo:



A isoquanta de $f(x,y) = \frac{x^2}{3} + \frac{y^2}{5}$ que passa pelo ponto $P(3,5)$ é $\frac{x^2}{3} + \frac{y^2}{5} = \frac{11}{2}$. Existem isoquantas de $f(x,y) = \frac{x^2}{3} + \frac{y^2}{5}$, de nível maior que $\frac{11}{2}$, que interceptam Ω , por exemplo, $\frac{x^2}{3} + \frac{y^2}{5} = 7$. Observe que o ângulo entre o vetor $\nabla f(3,5)$ e qualquer vetor $v \in \mathbb{R}^2$, tal que $\nabla g(3,5) \cdot v \geq 0$ e $\nabla h(3,5) \cdot v \geq 0$, é menor que 90° , portanto, se $x(t)$ é um caminho viável, $x(0) = (3,5)$, $x'(0) = v$, então $\nabla f(3,5) \cdot x'(0) > 0$ e essa última expressão está assegurando que $P(3,5)$ é um mínimo local.



Exemplo 4: Consideremos a generalização no $\mathbb{R}^n$:

$$\begin{aligned} \min \quad & f(x) \\ \text{sa} \quad & g(x^*) \cdot (x - x^*) \geq 0 \\ & h(x^*) \cdot (x - x^*) \leq 0 \end{aligned}$$

O cone $\Omega = \{x^* + t \cdot (x - x^*) \mid t \geq 0, \nabla g(x^*) \cdot (x - x^*) \geq 0, \nabla h(x^*) \cdot (x - x^*) \geq 0\}$ corresponde ao conjunto das semirretas, $t \geq 0$, com origem no ponto x^* na direção dos vetores $(x - x^*)$, tais que

$\nabla g(x^*) \cdot (x - x^*) \geq 0$ e $\nabla h(x^*) \cdot (x - x^*) \geq 0$. O vértice x^* do cone está na interseção das fronteiras $\nabla g(x^*) \cdot (x - x^*) = 0$ e $\nabla h(x^*) \cdot (x - x^*) = 0$.

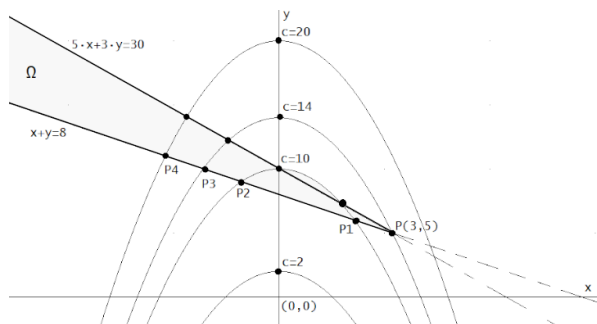
Suponhamos que x^* é um mínimo local para f na região Ω . Se a trajetória $x(t)$ é admissível, então os ângulos θ_f , θ_g e θ_h que o vetor velocidade $x'(0)$ faz com os gradientes de f , g e h , respectivamente, são tais que $\cos \theta_f \geq 0$, $\cos \theta_g \leq 0$ e $\cos \theta_h \geq 0$. Caso $\nabla f(x^*) \neq 0$ e $x'(0) \neq 0$, então $\|\nabla f(x^*)\| \cdot \cos \theta_f = \mu^* \cdot \|\nabla g(x^*)\| \cdot \cos \theta_g + \lambda^* \cdot \|\nabla h(x^*)\| \cdot \cos \theta_h$. Como podemos variar esses ângulos desde que $\cos \theta_f \geq 0$, $\cos \theta_g \leq 0$ e $\cos \theta_h \geq 0$, então $\lambda^* \leq 0$ e $\mu^* \geq 0$.

Exercício 1:

Considere o problema

$$\begin{aligned} \max \quad & x^2 + y \\ \text{sa} \quad & 5 \cdot x + 3 \cdot y \leq 30 \\ & x + y \geq 8. \end{aligned}$$

$\nabla f(3,5) = \begin{pmatrix} 6 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 3 \end{pmatrix} + \begin{pmatrix} -13 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \cdot \nabla g(3,5) + \mu \cdot \nabla h(3,5)$. Observando o desenho abaixo, convença-se que $P(3,5)$ é um máximo local:



Considere o problema:

$$\begin{aligned} \min \quad & x^2 + y \\ \text{sa} \quad & 5 \cdot x + 3 \cdot y \geq 30 \\ & x + y \leq 8. \end{aligned}$$

Observando o desenho acima, conclua que $P(3,5)$ é mínimo global.

Dada $f(x)$ sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_1 \geq 0, \dots, x_n \geq 0,$$

o lagrangeano dos problemas acima é:

$$\mathcal{L}(x, \lambda, \mu) = f(x) - \lambda_1 \cdot (g_1(x) - a_1) - \dots - \lambda_k \cdot (g_k(x) - a_k) - \mu_1 \cdot (h_1(x) - b_1) - \dots - \mu_l \cdot (h_l(x) - b_l) - \mu_{l+1} \cdot x_1 - \dots - \mu_{l+n} \cdot x_n.$$

As condições acima notadas por $KT1$ e $KT2$ podem ser expressas como condições sobre o lagrangeano:

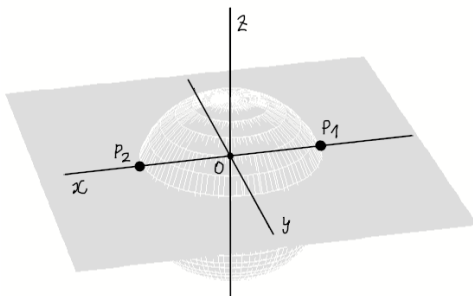
$$KT1: \frac{\partial \mathcal{L}}{\partial x_i}(x^*, \lambda^*, \mu^*) = \frac{\partial f}{\partial x_i}(x^*) - \lambda_1^* \cdot \frac{\partial g_1}{\partial x_i}(x^*) - \dots - \mu_l^* \cdot \frac{\partial h_l}{\partial x_i}(x^*) - \dots - \mu_{l+i}^* = 0;$$

$$KT2: \lambda_j^* \cdot \frac{\partial \mathcal{L}}{\partial \lambda_j}(x^*, \lambda^*, \mu^*) = 0, \dots, \mu_j^* \cdot \frac{\partial \mathcal{L}}{\partial \mu_j}(x^*, \lambda^*, \mu^*) = 0, \dots, \mu_{l+j}^* \cdot \frac{\partial \mathcal{L}}{\partial \mu_{l+j}}(x^*, \lambda^*, \mu^*) = 0, \dots$$

Exemplo 5: Considere o problema:

$$\begin{array}{ll} \min/\max & x \\ \text{sa} & (x-1)^2 + (y-1)^2 + (z-1)^2 \leq 9 \\ & z \geq 2. \end{array}$$

Observe o desenho abaixo. A região viável desse problema é a calota superior $z \geq 2$ da bola $(x-1)^2 + (y-1)^2 + (z-1)^2 \leq 9$ centrada no ponto $P(1,1,1)$. Não é difícil concluir que o mínimo da coordenada x está no ponto $P_1(1-2\sqrt{2}, 1, 2)$ e o máximo da coordenada x está no ponto $P_2(1+2\sqrt{2}, 1, 2)$. As coordenadas do ponto o são $O(1,1,2)$:



O lagrangeano desse problema é $\mathcal{L} = x - \lambda \cdot ((x-1)^2 + (y-1)^2 + (z-1)^2 - 9) - \mu \cdot (z-2)$. A primeira condição de K-T é:

$$\frac{\partial \mathcal{L}}{\partial x} = 1 - 2 \cdot \lambda \cdot (x-1) = 0$$

$$\frac{\partial \mathcal{L}}{\partial y} = -2 \cdot \lambda \cdot (y-1) = 0$$

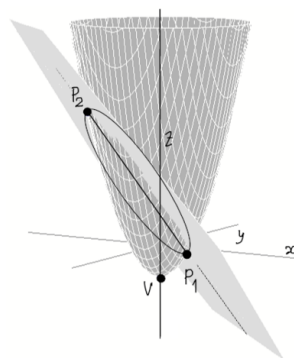
$$\frac{\partial \mathcal{L}}{\partial z} = -2 \cdot \lambda \cdot (z-1) - \mu = 0.$$

Como $y=1$ e $z=2$, então $\mu=-2\cdot\lambda$; como $x=1\pm 2\cdot\sqrt{2}$, então $\lambda=\pm\frac{1}{4\cdot\sqrt{2}}$. Considere o ponto $P_1(1-2\cdot\sqrt{2},1,2)$; nesse ponto $\lambda=-\frac{1}{4\cdot\sqrt{2}}$ e $\mu=\frac{1}{2\cdot\sqrt{2}}$, confirmando a terceira condição de K-T para um mínimo. Considere o ponto $P_2(1+2\cdot\sqrt{2},1,2)$; nesse ponto $\lambda=\frac{1}{4\cdot\sqrt{2}}$ e $\mu=-\frac{1}{2\cdot\sqrt{2}}$, confirmando a terceira condição de K-T para um máximo.

Exemplo 6: Considere o problema:

$$\begin{array}{ll} \min/\max & z \\ \text{s.a} & x^2 + y^2 - z \leq 1 \\ & x + y + z \leq 1. \end{array}$$

Observe o desenho abaixo. A região viável desse problema é a calota inferior do paraboloide $x^2 + y^2 - z \leq 1$ cortado pelo plano $x + y + z = 1$. O problema é determinar os valores mínimos e máximos para a coordenada z para pontos $P(x,y,z)$ na região viável.



Observando o desenho, concluímos que o ponto de máximo $P_2(x_2, y_2, z_2)$ está sobre a elipse determinada pela interseção do paraboloide e que o ponto de mínimo está no vértice $V(0,0,-1)$ do paraboloide.

O lagrangeano desse problema é $\mathcal{L} = z - \lambda_1 \cdot (x^2 + y^2 - z - 1) - \lambda_2 \cdot (x + y + z - 1)$. As condições de primeira ordem são:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= -2 \cdot \lambda_1 \cdot x - \lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= -2 \cdot \lambda_1 \cdot y - \lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial z} &= 1 + \lambda_1 - \lambda_2 = 0. \end{aligned}$$

No vértice $V(0,0,-1)$, temos $\lambda_2 = 0$, pois a fronteira $x+y+z-1$ não é ativa, portanto, a segunda condição de K-T para um mínimo assegura que $\lambda_2 \cdot (x+y+z-1) = 0$; como $\lambda_1 = -1$, então $V(0,0,-1)$ está satisfazendo as condições para um mínimo.

Em relação ao ponto de máximo $P_2(x_2, y_2, z_2)$, pelas condições de primeira ordem $x=y$ e como as duas restrições são ativas nesse ponto, $2 \cdot x^2 - 1 = z$ e $-2 \cdot x + 1 = z$ que resulta em $x^2 + x - 1 = 0$, cujas raízes são $x_1 = \frac{-1+\sqrt{5}}{2}$ e $x_2 = \frac{-1-\sqrt{5}}{2}$. Temos os pontos $P_1\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}, 2-\sqrt{5}\right)$ e $P_2\left(\frac{-1-\sqrt{5}}{2}, \frac{-1-\sqrt{5}}{2}, 2+\sqrt{5}\right)$ sobre as duas fronteiras. Os multiplicadores para $P_2\left(\frac{-1-\sqrt{5}}{2}, \frac{-1-\sqrt{5}}{2}, 2+\sqrt{5}\right)$ são $\lambda_1 = \frac{\sqrt{5}}{5}$ e $\lambda_2 = \frac{\sqrt{5}}{5} + 1$, o que corresponde a um máximo. Os multiplicadores para $P_1\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}, 2-\sqrt{5}\right)$ são $\lambda_1 = -\frac{\sqrt{5}}{5} < 0$ e $\lambda_2 = -\frac{\sqrt{5}}{5} + 1 > 0$. Esse ponto não pode ser nem um mínimo, nem um máximo e, de fato, não é; se nos deslocamos sobre a elipse, observamos que a coordenada z é mínima, mas se nos deslocamos a partir de $P_1\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}, 2-\sqrt{5}\right)$ sobre o parabolóide na direção do vértice $V(0,0,-1)$, a coordenada z diminui de valor.

Exemplo 7: Modifiquemos o problema anterior:

$$\begin{aligned} \min \quad & z \\ \text{s.a.} \quad & x^2 + y^2 - z \leq 1 \\ & x + y + z \geq 1. \end{aligned}$$

O lagrangeano desse problema é $\mathcal{L} = z - \lambda \cdot (x^2 + y^2 - z - 1) - \mu \cdot (x + y + z - 1)$. A primeira condição de K-T é:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= -2 \cdot \lambda \cdot x - \mu = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= -2 \cdot \lambda \cdot y - \mu = 0 \\ \frac{\partial \mathcal{L}}{\partial z} &= 1 + \lambda - \mu = 0. \end{aligned}$$

Nesse novo problema, temos dois pontos sobre as duas fronteiras: $P_1\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}, 2-\sqrt{5}\right)$, cujos multiplicadores são $\lambda = -\frac{\sqrt{5}}{5} < 0$ e $\mu = -\frac{\sqrt{5}}{5} + 1 > 0$; $P_2\left(\frac{-1-\sqrt{5}}{2}, \frac{-1-\sqrt{5}}{2}, 2+\sqrt{5}\right)$ são $\lambda = \frac{\sqrt{5}}{5}$ e $\mu = \frac{\sqrt{5}}{5} + 1$. Portanto, $P_1\left(\frac{-1+\sqrt{5}}{2}, \frac{-1+\sqrt{5}}{2}, 2-\sqrt{5}\right)$ está satisfazendo a segunda e a terceira condições de K-T para um mínimo e, de fato, é o mínimo. Esse problema não tem máximos. Por sua vez, $P_2\left(\frac{-1-\sqrt{5}}{2}, \frac{-1-\sqrt{5}}{2}, 2+\sqrt{5}\right)$ não é nem um mínimo, nem um máximo.

Exemplo 8: Considere o problema:

$$\begin{array}{ll} \min & x^2 + (y-1)^2 \\ \text{sa} & x^2 - y \leq 0 \end{array}$$

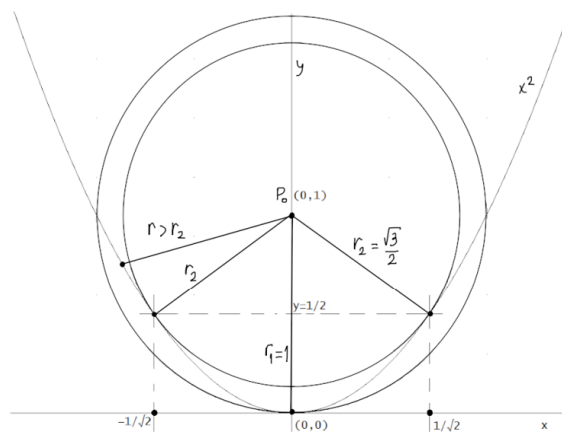
O lagrangeano desse problema é $\mathcal{L} = x^2 + (y-1)^2 - \lambda \cdot (x^2 - y)$. A primeira condição de K-T é:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 2 \cdot x \cdot (1 - \lambda) = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 2 \cdot (y-1) + \lambda = 0. \end{aligned}$$

Uma solução para esse problema é $P_0(0,1)$, que obviamente é o ponto de mínimo global e o ponto crítico da função $f(x,y) = x^2 + (y-1)^2$, como a fronteira $x^2 - y \leq 0$ é inativa $\lambda = 0$.

A origem $(0,0)$ satisfaz as três condições de K-T com $\lambda = 2$ para um máximo local. De fato, $(0,0)$ é um máximo local: $r_1 = 1$.

Outros dois pontos, $(\pm \frac{1}{\sqrt{2}}, \frac{1}{2})$, estão sobre a circunferência $r_2 = \frac{\sqrt{3}}{2}$ e $\lambda = 1$ e satisfazem as condições de K-T para máximo, mas não são nem mínimos, nem máximos!



Portanto, as condições de K-T são condições necessárias, mas não suficientes!

Exercício 2

Classifique os pontos críticos do lagrangeano utilizando o teorema de Kuhn-Tucker:

i) $f(x, y, z) = x^2 + y^2 + z^2 - x \cdot y - 3 \cdot y$, sujeita à restrição $g(x, y, z) = z - x - y \leq 1$;

ii) $f(x, y) = 2 \cdot x^3 + x \cdot y^2 + y^2$, sujeita à restrição $g(x, y) = x^2 + y^2 \leq 1$;

iii) $f(x, y, z) = x^2 + 2 \cdot y^2 + z^2 - x \cdot y - 2 \cdot z$, sujeita à restrição $h(x, y, z) = x + 3 \cdot y + 2 \cdot z \geq 8$;

iv) $f(x, y, z) = 5 \cdot x^2 + 10 \cdot y^2 + z^2 - 4 \cdot x \cdot y - 2 \cdot x \cdot z - 36 \cdot y$, sujeita à restrição $h(x, y, z) = x + y + z \geq 3$.

Exercício 3

Considere o problema $f(x, y, z) = x \cdot y$, sujeita às restrições $x^2 + y^2 + z^2 \leq 6$ e $x + y - z \geq 0$.

i) Escreva as condições de Kuhn-Tucker para mínimo local e máximo local do problema acima;

ii) Usando as condições de Kuhn-Tucker, classifique, ou seja, indique a possibilidade de ser mínimo ou máximo local, os pontos críticos do lagrangeano: $(1, 1, 2)$ e $(-1, -1, -2)$.

Exercício 4

Escreva as condições de Kuhn-Tucker para um mínimo local do problema $f(x, y, z) = x \cdot y \cdot z$, sujeita à restrição $x + y - z \leq 0$ e $x^2 + y^2 + z^2 \leq 9$.

Exercício 5

Mostre que $(1, 2)$ satisfaz as condições de Kuhn-Tucker para um mínimo local do problema $2 \cdot x^2 + 2 \cdot x \cdot y + y^2 - 10 \cdot x - 10 \cdot y$, sujeita à restrição $x^2 + y^2 \leq 5$, $3 \cdot x + y \leq 6$, $x \geq 0$, $y \geq 0$.

Exercício 6

Considere o ponto $P(1, 2, 2)$ na interseção das duas regiões:

$$x + y + 2 \cdot z \geq 7 \text{ e } 2 \cdot x + 2 \cdot y + \frac{z^2}{2} \leq 8.$$

Utilizando as condições de Kuhn-Tucker, justifique a condição de $P(1, 2, 2)$ ser um mínimo, máximo local ou nem mínimo, nem máximo para:

$$f(x, y, z) = x + 2 \cdot y + 2 \cdot z - x \cdot y - y \cdot z,$$

em relação às restrições acima.

Exercício 7

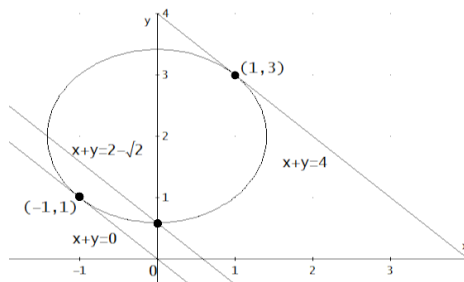
Encontre o ponto de máximo global e o valor máximo global de $x_1 \cdot x_2$ sujeita à restrição:

$$x_1 + x_2 \leq 10,$$

$$x_1 + 2 \cdot x_2 \leq 18$$

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Exercício 8: Determine o ponto de mínimo e máximo global de $f(x, y) = x + y$, sujeita às restrições $x^2 + (y - 2)^2 = 2$, $x \geq 0$, $y \geq 0$. Observe o desenho abaixo:



Exercício 9

Escreva as condições de Kuhn-Tucker para um máximo local do problema $f(x, y, z) = 4z - x^2 - y^2 - z^2$, sujeita à restrição $z - x \cdot y \geq 0$ e $x^2 + y^2 + z^2 \leq 3$.

Exercício 10

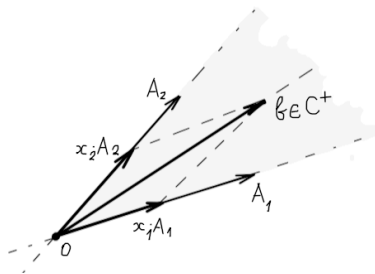
Encontre o valor mínimo e o valor máximo de $2 \cdot x^3 + x \cdot y^2 + y^2$, sujeita a restrição $x^2 + y^2 \leq 1$.

4. Lema de Farkas

Gyula Farkas (1847-1930) foi um matemático húngaro que publicou, em 1901, o resultado de seus estudos sobre sistemas mecânicos, chamado “**Lema de Farkas**”⁶. O resultado, após a publicação do artigo “*Nonlinear Programming*”, de Kuhn e Tucker (1951), tornou-se fundamental para a teoria das desigualdades lineares. Mas, esse resultado já estava sendo empregado antes de 1950, pois Dantzig afirma que von Neumann apresentou-lhe o “Lema de Farkas” no contexto de uma conversa sobre programação linear em 1947. Além disso, o Lema já havia sido utilizado na dissertação de mestrado de Karush em 1939.

Cones – Um cone $C \subseteq \mathbb{R}^m$ é um subconjunto tal que se $v_1, v_2 \in C \rightarrow v_1 + v_2 \in C$ e $\lambda \geq 0, v \in C \rightarrow \lambda \cdot v \in C$.

Consideremos os vetores $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} = A_1, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = A_n \in \mathbb{R}^m$. O cone positivo $C^+ \subseteq \mathbb{R}^m$ gerado pelos vetores $A_1, \dots, A_n$ é o conjunto de todas as combinações lineares não negativas desses vetores: $b = x_1 \cdot A_1 + \dots + x_n \cdot A_n = x_1 \cdot \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + x_n \cdot \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}, x_1 \geq 0, \dots, x_n \geq 0$.

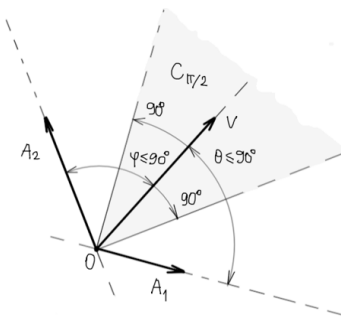


Se considerarmos a matriz $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$, cujas colunas são os vetores $A_1, \dots, A_n$, então

$$\begin{aligned} A \cdot x &= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n \\ \vdots \\ a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n \end{pmatrix} \\ &= x_1 \cdot \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + x_n \cdot \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = x_1 \cdot A_1 + \dots + x_n \cdot A_n = b \end{aligned}$$

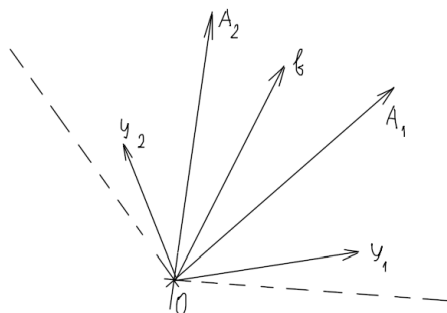
Portanto, o cone positivo gerado por esses vetores é o conjunto $C^+(A_1, \dots, A_n) = \{A \cdot x \in \mathbb{R}^m \mid x \geq 0, x \in \mathbb{R}^n\}$. Observe que $C^+(A_1, \dots, A_n) \subseteq \mathbb{R}^m$.

Novamente, consideremos os vetores $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} = A_1, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = A_n \in \mathbb{R}^m$. O cone $C_{\pi/2}$ é o conjunto dos vetores $v = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{R}^m$ que não fazem ângulos maiores que 90° com os vetores $A_1, \dots, A_n$, isto é, $v \cdot A_1 \geq 0, v \cdot A_2 \geq 0, \dots, v \cdot A_n \geq 0$. Se considerarmos a matriz $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$, cujas colunas são os vetores $A_1, \dots, A_n$, então $v \cdot A = (x_1 \ \dots \ x_m) \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \geq 0$. Portanto, $C_{\pi/2}$ é o conjunto dos vetores $C_{\pi/2}(A_1, \dots, A_n) = \{x \in \mathbb{R}^m \mid x' \cdot A \geq 0\}$. Observe que $C_{\pi/2}(A_1, \dots, A_n) \subseteq \mathbb{R}^m$.

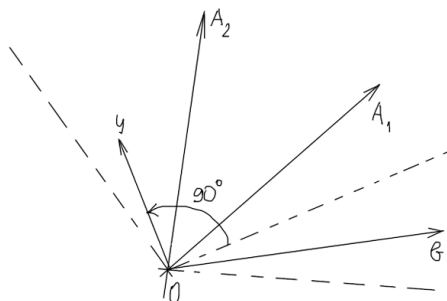


No desenho acima, θ é ângulo entre o vetor v e o vetor A_1 e φ é ângulo entre o vetor v e o vetor A_2 . O cone $C_{\pi/2}(A_1, A_2)$ é o conjunto dos vetores v , tais que esses ângulos sejam de, no máximo, 90° .

O Lema de Farkas afirma algo intuitivamente óbvio: um vetor $b \in \mathbb{R}^m$, $b \neq 0$ está no cone positivo $C^+(A_1, \dots, A_n) = \{A \cdot x \mid x \geq 0\}$, gerado por $A_1, \dots, A_n$ se, e somente se, todo o vetor $y \in \mathbb{R}^m$ que mantém ângulos menores ou iguais a 90° com cada um dos vetores $A_1, \dots, A_n$, isto é, $y' \cdot A_i \geq 0$, também faz um ângulo menor ou igual a 90° com b , isto é, $y' \cdot b \geq 0$. Expressando em termos de cones: $b \neq 0$, $b \in C^+(A_1, \dots, A_n) \Leftrightarrow \forall y, y \in C_{\pi/2}(A_1, \dots, A_n) \rightarrow y \cdot b \geq 0$. Entretanto, a melhor maneira de expressar o lema de Farkas é em termos matriciais: $\forall b \neq 0, A \cdot x = b, x \geq 0$ se, e só se, $\forall y, y' \cdot A \geq 0$, então $y \cdot b \geq 0$.



Equivalentemente, um vetor $b \neq 0$ não pertence ao cone positivo $C^+(A_1, \dots, A_n) = \{A \cdot x \mid x \geq 0\}$ se, e somente se, existe um vetor $y \in \mathbb{R}^m$ que mantém ângulos menores ou iguais a 90° com cada um dos vetores $A_1, \dots, A_n$, isto é, $y' \cdot A_i \geq 0$, e faz um ângulo maior que 90° com b , $y' \cdot b < 0$. Veja o desenho abaixo:



Ainda que a afirmação do Lema de Farkas seja facilmente compreendida, a demonstração formal apresenta certas dificuldades e se estende para além do nosso objetivo.

Normalmente, o Lema de Farkas é apresentado numa versão matricial muito adequada para uso em demonstrações, mas que encobre o conteúdo intuitivo da afirmação. A versão a seguir será a que empregaremos.

Lema de Farkas: Seja A matriz $m \times n$, $b \neq 0$, então o sistema $A \cdot x = b$ tem solução $x \geq 0$ se, e somente se, para todo $y \in \mathbb{R}^m$, tal que $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$.

5. A demonstração I do Teorema de Kuhn-Tucker

Primeiramente, demonstraremos o teorema para a situação em que x^* é um máximo local para $f(x)$ e temos apenas duas fronteiras ativas, $g_1(x) \leq a_1$ e $h_1(x) \geq b_1$, e

as demais inativas. A situação geral não é diferente dessa situação particular, exceto pela notação.

Seja $x: [0, +\infty) \rightarrow \Omega$ uma **trajetória admissível** em x^* , ou seja, $x(0) = x^*$, $g_1(x(0)) = g_1(x^*) = a_1$, $h_1(x(0)) = h_1(x^*) = b_1$ e $\left. \frac{d}{dt} g_1(x(t)) \right|_{t=0} = \nabla g_1(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla g_1(x^*) \cdot x'(0) \leq 0$, $\left. \frac{d}{dt} h_1(x(t)) \right|_{t=0} = \nabla h_1(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla h_1(x^*) \cdot x'(0) \geq 0$.

Portanto,

$$(x'_1(0) \ \cdots \ x'_n(0)) \cdot \begin{pmatrix} -\frac{\partial g_1}{\partial x_1}(x^*) & \frac{\partial h_1}{\partial x_1}(x^*) \\ \vdots & \vdots \\ -\frac{\partial g_1}{\partial x_n}(x^*) & \frac{\partial h_1}{\partial x_n}(x^*) \end{pmatrix} = (-\nabla g_1(x^*) \cdot x'(0) \ \nabla h_1(x^*) \cdot x'(0)) \geq 0,$$

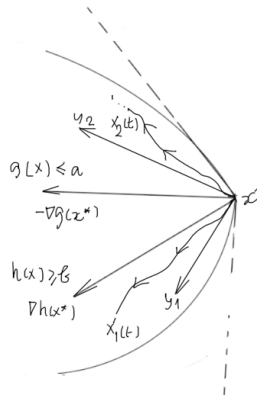
como x^* é um máximo local para f , então $\left. \frac{d}{dt} f(x(t)) \right|_{t=0} = \nabla f(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla f(x^*) \cdot x'(0) \leq 0$, invertendo o sinal, $\left. -\frac{d}{dt} f(x(t)) \right|_{t=0} = -\nabla f(x(t)) \cdot x'(t) \Big|_{t=0} = -\nabla f(x^*) \cdot x'(0) \geq 0$.

No **Lema de Farkas**, fazemos $x'(0) = y$,

$$\begin{pmatrix} -\frac{\partial g_1}{\partial x_1}(x^*) & \frac{\partial h_1}{\partial x_1}(x^*) \\ \vdots & \vdots \\ -\frac{\partial g_1}{\partial x_n}(x^*) & \frac{\partial h_1}{\partial x_n}(x^*) \end{pmatrix} = A$$

e $-\nabla f(x^*) = b$, então obtemos que $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$.

Para que possamos aplicar o Lema de Farkas, temos que verificar a proposição $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$ para todo $y \in \mathbb{R}^n$, tal que $y' \cdot A \geq 0$, logo, temos que assumir uma **hipótese adicional**: para todo $y \in \mathbb{R}^n$, tal que $y' \cdot A \geq 0$, existe uma **trajetória admissível** $x(t)$ em x^* , tal que $x'(0) = y$. Essa **hipótese adicional** chamamos de “**qualificação de restrição**”. Veja o desenho a seguir. Em outras palavras, $\forall y \in \mathbb{R}^n, \nabla g_1(x^*) \cdot y \leq 0, \nabla h_1(x^*) \cdot y \geq 0, \exists x(t), x(0) = x^* \rightarrow x'(0) = y$:



Admitida a hipótese acima, pelo Lema de Farkas existe $\begin{pmatrix} \lambda_1 \\ -\mu_1 \end{pmatrix} \geq 0$, tal que

$$A \cdot \begin{pmatrix} \lambda_1 \\ -\mu_1 \end{pmatrix} = \begin{pmatrix} -\frac{\partial g_1}{\partial x_1}(x^*) & \frac{\partial h_1}{\partial x_1}(x^*) \\ \vdots & \vdots \\ -\frac{\partial g_1}{\partial x_n}(x^*) & \frac{\partial h_1}{\partial x_n}(x^*) \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ -\mu_1 \end{pmatrix} = b = \begin{pmatrix} -\frac{\partial f}{\partial x_1}(x^*) \\ \vdots \\ -\frac{\partial f}{\partial x_n}(x^*) \end{pmatrix}.$$

Isto é, $-\lambda_1 \cdot \nabla g_1(x^*) + (-\mu_1) \cdot \nabla h_1(x^*) = -\nabla f(x^*)$, ainda,

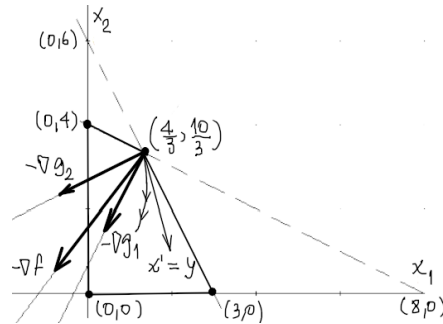
$$\lambda_1 \cdot \nabla g_1(x^*) + \mu_1 \cdot \nabla h_1(x^*) = \nabla f(x^*),$$

$\lambda_1 \geq 0$ e $\mu_1 \leq 0$. Os demais multiplicadores, que correspondem às fronteiras inativas, são nulos. Portanto, acabamos de demonstrar o Teorema de Kuhn-Tucker.

Exemplo 9: Tudo o que estamos estudando de Programação Não-Linear aplica-se à Programação Linear como um caso particular. Vejamos um exemplo:

$$\begin{aligned} \max \quad & 3 \cdot x_1 + 4 \cdot x_2 \\ \text{sa} \quad & x_1 + 2 \cdot x_2 \leq 8 \\ & 2 \cdot x_1 + x_2 \leq 6 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

O ponto ótimo está no vértice $(x^*, y^*) = (4/3, 10/3)$.

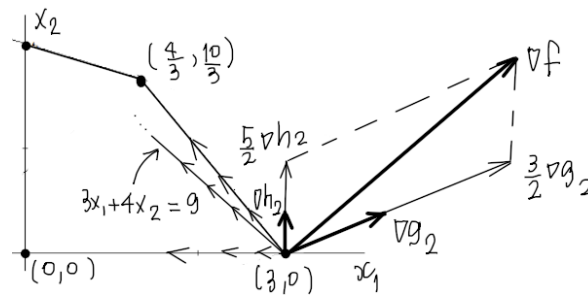


Observe que, para qualquer vetor y , $y' \cdot (-\nabla g_1(x^*, y')) \geq 0$ e $y' \cdot (-\nabla g_2(x^*, y')) \geq 0$, existe uma curva admissível $x(t)$, tal que $x'(0) = y$ e $y' \cdot (-\nabla f) \geq 0$; $y' \cdot (-\nabla f) \geq 0$. Pelo Lema de Farkas, existem $\lambda_1 \geq 0$ e $\lambda_2 \geq 0$, tais que $-\nabla f = \lambda_1 \cdot (-\nabla g_1) + \lambda_2 \cdot (-\nabla g_2)$, isto é, $\nabla f = \lambda_1 \cdot \nabla g_1 + \lambda_2 \cdot \nabla g_2$, ainda,

$$\nabla f = \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{5}{3} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{2}{3} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \lambda_1 \cdot \nabla g_1 + \lambda_2 \cdot \nabla g_2.$$

Nesse problema existem duas fronteiras adicionais: $x_1 \geq 0, x_2 \geq 0$; ambas são inativas no caso do vértice $(4/3, 10/3)$, portanto, os correspondentes multiplicadores μ_1, μ_2 são nulos.

Veamos a situação do vértice $(3,0)$, onde a fronteira $x_2 = 0$ é ativa; a outra fronteira ativa é $2 \cdot x_1 + x_2 = 6$. Observe que o vértice $(3,0)$ satisfaz a restrição de qualificação.



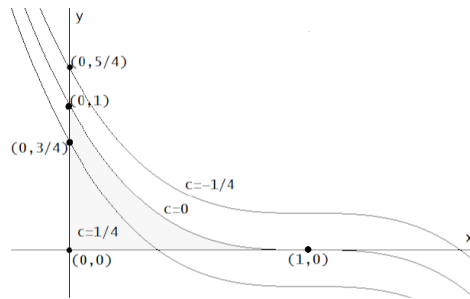
$$\nabla f = \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{3}{2} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \frac{5}{2} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \lambda_2 \cdot \nabla g_2 + \mu_2 \cdot \nabla h_2.$$

O vértice $(3,0)$ não está satisfazendo as condições nem de mínimo e nem de máximo e, de fato, não pode satisfazer. Se nos deslocamos a partir do ponto $(3,0)$ na direção da origem, estamos fazendo um ângulo maior que 90° , com $\nabla f = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ e, portanto, f diminui; se nos deslocamos sobre o segmento que une os

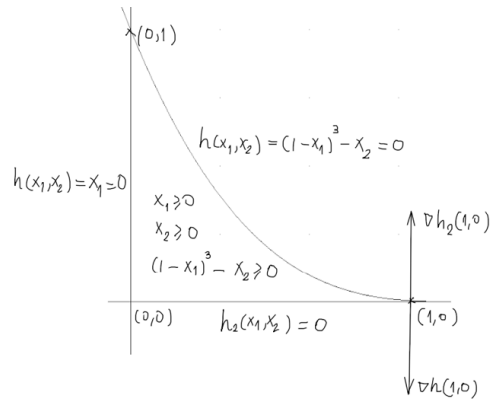
vértices $(3,0)$ e $(4/3,10/3)$ partindo de $(3,0)$, f aumenta, pois o ângulo entre $\nabla f = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ e $(-5/3, 10/3) = (4/3, 10/3) - (3,0)$, que está na direção da reta, é menor que 90° ; basta fazer o produto escalar desses vetores para verificar essa última afirmação. Se nos deslocamos ao longo da isoquanta $3 \cdot x_1 + 4 \cdot x_2 = 9$ partindo de $(3,0)$, f permanece constante: $f(x_1, x_2) = 3 \cdot x_1 + 4 \cdot x_2 = 9$.

Nos exemplos 3 e 4, os conjuntos Ω determinados pelas restrições são os cones $C_{\pi/2}(-\nabla g(x^*), \nabla h(x^*))$ e $C_{\pi/2}(\nabla g(x^*), \nabla h(x^*))$, respectivamente, então a qualificação de restrição é automaticamente satisfeita. O mesmo acontece nos problemas lineares, como no exemplo 9.

Exemplo 10 – Exemplo de Topkins: Kuhn e Tucker fornecem um exemplo no qual a qualificação de restrição não é satisfeita. Considere a região Ω , esboçada abaixo, definida pelas restrições $h(x_1, x_2) = (1-x_1)^3 - x_2 \geq 0$, $h_1(x_1, x_2) = x_1 \geq 0$ e $h_2(x_1, x_2) = x_2 \geq 0$. No desenho abaixo, representamos as isoquantas $h(x_1, x_2) = (1-x_1)^3 - x_2 = c$ para $c = -\frac{1}{4}, 0, \frac{1}{4}$. A região cinza é Ω :



Observe que $\Delta h(1,0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ e $\Delta h_2(1,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. As fronteiras ativas são $h_2(x_1, x_2) = x_2 \geq 0$ e $h(x_1, x_2) = (1-x_1)^3 - x_2 \geq 0$. Se $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ é tal que $y' \cdot \Delta h_2(1,0) = y_2 \geq 0$ e $y' \cdot \Delta h(1,0) = -y_2 \geq 0$, então $y_2 = 0$. Portanto, $y = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}$, ou seja, o cone $C_{\pi/2}(\nabla h(1,0), \nabla h_2(1,0)) = \{(y_1, 0) | y_1 \in \mathbb{R}\}$. Entretanto, as trajetórias admissíveis $x: [0, +\infty) \rightarrow \Omega$, $x(0) = (1,0)$ são tais que, necessariamente, $x'(0) = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}$, $y_1 \leq 0$. Portanto, por exemplo, não existe trajetória admissível $x: [0, +\infty) \rightarrow \Omega$, tal que $x'(0) = (1,0) = e_1$, no entanto, $e_1' \cdot \Delta h_2(1,0) = e_1' \cdot \Delta h(1,0) \geq 0$. Sendo assim, a qualificação de restrição não é satisfeita.



Exemplo 11: Considere o problema:

$$\begin{array}{ll} \max & x_1 + x_2 \\ \text{sa} & h(x_1, x_2) = (1-x_1)^2 - x_2 \geq 0 \\ & h_1(x_1, x_2) = x_1 \geq 0 \\ & h_2(x_1, x_2) = x_2 \geq 0. \end{array}$$

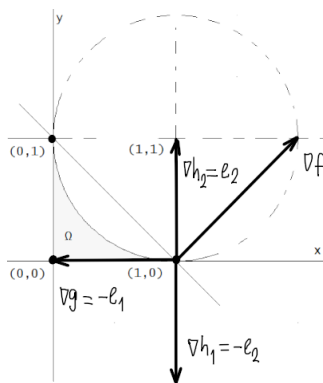
Nesse problema, a primeira condição de Kuhn-Tucker não é satisfeita. O ponto $(1,0)$ é um máximo global de $f(x_1, x_2) = x_1 + x_2$, mas $\nabla f(1,0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ não pode ser expresso como combinação de $\Delta h(1,0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ e $\Delta h_2(1,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Kuhn escreveu que, numa apresentação preliminar do artigo “*Nonlinear Programming*”, foi-lhes chamada a atenção para o exemplo acima e que isso exigiu o acréscimo da hipótese da qualificação de restrição⁷. A hipótese foi introduzida de maneira simples, permitindo o emprego do Lema de Farkas e contornando o problema. Kuhn observa que a condição necessária para o uso do Lema de Farkas é satisfeita no caso dos problemas lineares e que o Teorema de Kuhn-Tucker resulta da “linearização” dos problemas não-lineares (através dos gradientes), mas obviamente a extensão para os problemas não-lineares não pode ser feita sem alguma condição adicional⁸. Repetimos: essa condição é a hipótese de “qualificação de restrição”!

Exemplo 12: Considere o problema:

$$\begin{array}{ll} \max & x + y \\ \text{sa} & x \leq 1 \\ & (x-1)^2 + (y-1)^2 \geq 1 \\ & x \geq 0, y \geq 0. \end{array}$$

A região Ω está entre as retas $x=0$, $x=1$, $y=0$ e parte exterior do círculo $(x-1)^2 + (y-1)^2 = 1$ limitada pelas retas. Não é difícil ver que o ponto $(1,0)$ é o máximo desse problema. Em $(1,0)$, as fronteiras $g(x,y)=x=1$, $h_1(x,y)=(x-1)^2 + (y-1)^2 = 1$ e $h_2(x,y)=y=0$ são ativas. As fronteiras $y=0$ e $(x-1)^2 + (y-1)^2 = 1$ se tangenciam em $(1,0)$.



No ponto $(1,0)$ a qualificação de restrição é satisfeita. O cone $C_{\pi/2}(-\nabla g(1,0), \nabla h_1(1,0), \nabla h_2(1,0)) = C_{\pi/2}(-e_1, e_2, -e_2) = \{(\theta, 0) \mid \theta \leq 0\}$, pois os caminhos $x(t) = (1 + \theta \cdot t, 0)$, $\theta < 0$ são viáveis: $x(0) = (1,0)$ e $x'(0) = (\theta, 0)$.

Observe que $\nabla f(1,0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 1 \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \lambda \cdot \nabla g(1,0) + \mu_1 \cdot \nabla h_1(1,0) + \mu_2 \cdot \nabla h_2(1,0)$.

Portanto, as condições de Kuhn-Tucker para um máximo estão sendo satisfeitas no ponto $(1,0)$.

A introdução *ad hoc* da qualificação de restrição por parte dos autores ensejou a busca por condições que permitissem a demonstração do Teorema de Kuhn-Tucker e fossem mais evidentes em relação aos problemas comuns da análise econômica. A qualificação de restrição, como foi apresentada por Kuhn e Tucker, tem uma falsa aparência de simplicidade, como pudemos suspeitar no problema anterior, que geometricamente é parecido com o exemplo de Topkins.

No artigo “*Lagrange Multipliers Revisited*”, de 1950⁹, escrito por Morton Slater, aparece uma condição que passou a ser denominada “condição de Slater”. O objetivo de Slater não era obter alternativas para a qualificação de restrição, mas retirar a necessidade de diferenciabilidade que Kuhn e

Tucker impuseram às funções. A condição de Slater surge como um subproduto do trabalho.

Outros artigos bem conhecidos são o “*Constraint Qualifications in Maximization Problems*”, de 1961¹⁰, e o “*Constraint Qualifications in Maximization Problems II*”, de 1960¹¹, nos quais os autores simplificam os resultados do trabalho anterior. As datas das publicações não correspondem à ordem cronológica da elaboração dos trabalhos.

Com o transcorrer das análises, passamos a entender por “restrição de qualificação” qualquer hipótese que seja suficiente para assegurar as conclusões necessárias do Teorema de Kuhn-Tucker. Cada uma dessas hipóteses foi elaborada tendo em vista o tipo de problema sob análise. Isso gerou uma grande variedade de qualificações de restrições, todas assegurando as conclusões do Teorema de Kuhn-Tucker, mas não logicamente equivalentes, e gerando uma estrutura muito complexa de relações lógicas. Em 1973, um autor chamado David Peterson empreendeu uma análise das relações num *paper* surpreendente e intrincado¹².

Nós voltaremos às qualificações de restrições na próxima Lição. Não devemos nos preocupar muito com a qualificação da restrição em relação aos problemas que nos interessam. Todos satisfazem essa condição. Por enquanto, apenas sublinhamos que os problemas da programação linear satisfazem a qualificação de restrição tal como foi formulada por Kuhn e Tucker.

6. Demonstração II do Teorema de Kuhn-Tucker

A seguir, simplesmente repetimos a argumentação anterior com uma notação mais geral. Suponhamos que x^* seja um máximo local para $f(x)$ sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_1 \geq 0, \dots, x_n \geq 0.$$

O lagrangeano desse problema é:

$$\begin{aligned}\mathcal{L}(x, \lambda, \mu) = & f(x) - \lambda_1 \cdot (g_1(x) - a_1) - \dots - \lambda_k \cdot (g_k(x) - a_k) - \\ & \mu_1 \cdot (h_1(x) - b_1) - \dots - \mu_l \cdot (h_l(x) - b_l) - \\ & \mu_{l+1} \cdot x_1 - \dots - \mu_{l+n} \cdot x_n\end{aligned}$$

Digamos que as fronteiras $g_1(x) = a_1, \dots, g_k(x) = a_k, h_1(x) = b_1, \dots, h_l(x) = b_l, x_{n_1} = 0, \dots, x_{n_p} = 0$ sejam ativas e que $x(t)$ seja uma curva admissível em x^* . Então,

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_k}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_l}{\partial x_1}(x^*) & \dots & \frac{\partial g_l}{\partial x_n}(x^*) \\ \frac{\partial h_1}{\partial x_1}(x^*) & \dots & \frac{\partial h_1}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_p}{\partial x_1}(x^*) & \dots & \frac{\partial h_p}{\partial x_n}(x^*) \\ \delta_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \delta_n \end{pmatrix} \cdot \begin{pmatrix} x'_1(0) \\ \vdots \\ x'_n(0) \end{pmatrix} = \begin{pmatrix} -\nabla g_1(x^*) \cdot x'(0) \\ \vdots \\ -\nabla g_l(x^*) \cdot x'(0) \\ \nabla h_1(x^*) \cdot x'(0) \\ \vdots \\ \nabla h_p(x^*) \cdot x'(0) \\ \delta_1 \cdot x'_1(0) \\ \vdots \\ \delta_n \cdot x'_n(0) \end{pmatrix} \geq 0.$$

δ_k é definido da seguinte maneira:

$$\delta_k = \begin{cases} 1, k \in \{n_1, n_2, \dots, n_p\}; \\ 0, k \notin \{n_1, n_2, \dots, n_p\}. \end{cases}$$

O resultado acima, obviamente, resulta do fato que $x(t)$ é uma curva admissível em x^* .

Uma vez mais, o Lema de Farkas: **se para todo** $y \in \mathbb{R}^m$ que satisfaz $y' \cdot A \geq 0$, então $y' \cdot b \geq 0$, concluímos que $A \cdot x = b$ tem solução $x \geq 0$. No caso, $x'(0)$ faz o papel de y e devemos ter o cuidado de que os vetores tangentes $x'(0)$ desempenhem realmente esse papel, ou seja, que para **todo** $y \in \mathbb{R}^m$, tal que $y' \cdot A \geq 0$, exista uma curva admissível $x: [0, +\infty) \rightarrow \Omega$, tal que $x'(0) = y$. Essa exigência é a qualificação de restrição. Para fazer o papel de b , escolhemos $-\nabla f(x^*)$. Portanto, devemos exigir que $x^* \in \Omega$ satisfaça a condição adicional de qualificação de restrição, isto é, que se as fronteiras $g_1(x) = a_1, \dots, g_k(x) = a_k, h_1(x) = b_1, \dots, h_l(x) = b_l, x_{n_1} = 0, \dots, x_{n_p} = 0$ são ativas e $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ é tal que:

$$(y_1 \quad \dots \quad y_n) \cdot \begin{pmatrix} -\frac{\partial g_{i_1}}{\partial x_1}(x^*) & \dots & -\frac{\partial g_{i_r}}{\partial x_1}(x^*) & \frac{\partial h_{j_1}}{\partial x_1}(x^*) & \dots & \frac{\partial h_{j_s}}{\partial x_1}(x^*) & \delta_{n_1} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial g_{i_1}}{\partial x_n}(x^*) & \dots & -\frac{\partial g_{i_r}}{\partial x_n}(x^*) & \frac{\partial h_{j_1}}{\partial x_n}(x^*) & \dots & \frac{\partial h_{j_s}}{\partial x_n}(x^*) & 0 & \dots & \delta_{n_p} \end{pmatrix} \geq 0,$$

então exista $x(t)$, curva admissível em x^* , tal que $\begin{pmatrix} x'_1(0) \\ \vdots \\ x'_n(0) \end{pmatrix} = y$.

Enfatizamos que o ponto x^* **satisfaz a qualificação de restrição**. Como x^* é um máximo local para $f(x)$, então

$$-\frac{d}{dt}f(x(t))\Big|_{t=0} = -\nabla f(x(t)) \cdot x'(t)\Big|_{t=0} = -\nabla f(x^*) \cdot x'(0) \geq 0.$$

Então, pelo Lema de Farkas, existe $\begin{pmatrix} \lambda_h^* \\ \vdots \\ \lambda_r^* \\ -\mu_{j_1}^* \\ \vdots \\ -\mu_{j_s}^* \\ -\nu_1^* \\ \vdots \\ -\nu_n^* \end{pmatrix} \geq 0$, tal que

$$\begin{pmatrix} -\frac{\partial g_{i_1}}{\partial x_1}(x^*) & \dots & -\frac{\partial g_{i_r}}{\partial x_1}(x^*) & \frac{\partial h_{j_1}}{\partial x_1}(x^*) & \dots & \frac{\partial h_{j_s}}{\partial x_1}(x^*) & \delta_1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial g_{i_1}}{\partial x_n}(x^*) & \dots & -\frac{\partial g_{i_r}}{\partial x_n}(x^*) & \frac{\partial h_{j_1}}{\partial x_n}(x^*) & \dots & \frac{\partial h_{j_s}}{\partial x_n}(x^*) & 0 & \dots & \delta_n \end{pmatrix} \cdot \begin{pmatrix} \lambda_h^* \\ \vdots \\ \lambda_r^* \\ -\mu_{j_1}^* \\ \vdots \\ -\mu_{j_s}^* \\ -\nu_1^* \\ \vdots \\ -\nu_n^* \end{pmatrix} = \begin{pmatrix} -\frac{\partial f}{\partial x_1}(x^*) \\ \vdots \\ -\frac{\partial f}{\partial x_n}(x^*) \end{pmatrix}.$$

Se $\nabla f(x^*)=0$, obviamente, $\lambda_h^* = \dots = \lambda_r^* = \mu_{j_1}^* = \dots = \mu_{j_s}^* = \nu_1^* = \dots = \nu_n^* = 0$. Portanto,

$$\begin{aligned} \nabla f(x^*) = & \lambda_h^* \cdot \nabla g_{i_1}(x^*) + \dots + \lambda_r^* \cdot \nabla g_{i_r}(x^*) + \\ & \mu_{j_1}^* \cdot \nabla h_{j_1}(x^*) + \dots + \mu_{j_s}^* \cdot \nabla h_{j_s}(x^*) + \\ & \delta_1 \cdot \nu_1^* \cdot \tilde{e}_1 + \dots + \delta_n \cdot \nu_n^* \cdot \tilde{e}_n. \end{aligned}$$

Se as fronteiras $g_i(x)=a_i$ e $h_j(x)=b_j$ são inativas em x^* , ou seja, $g_i(x^*) < a_i$ e $h_j(x^*) > b_j$, então os correspondentes multiplicadores são nulos, isto é, $\lambda^* = 0$ e $\mu^* = 0$. Portanto, o gradiente $\nabla f(x^*)$ é uma combinação linear dos gradientes das funções que definem as restrições:

$$\nabla g_1(x^*), \dots, \nabla g_k(x^*), \nabla h_1(x^*), \dots, \nabla h_l(x^*), \nabla \pi_1(x^*), \dots, \nabla \pi_n(x^*);$$

os correspondentes multiplicadores são tais que $\lambda_1^* \geq 0, \dots, \lambda_k^* \geq 0, \mu_1^* \leq 0, \dots, \mu_l^* \leq 0, \nu_1^* \leq 0, \dots, \nu_n^* \leq 0$.

Se determinada fronteira é inativa, o correspondente multiplicador é necessariamente nulo. Expressamos essa condição da seguinte maneira:

$$\lambda_1^* \cdot (g_1(x^*) - a_1) = 0, \dots, \lambda_k^* \cdot (g_k(x^*) - a_k) = 0,$$

$$\mu_1^* \cdot (h_1(x^*) - b_1) = 0, \dots, \mu_l^* \cdot (h_l(x^*) - b_l) = 0,$$

$$\nu_1^* \cdot x_1^* = 0, \dots, \nu_n^* \cdot x_n^* = 0.$$

As três condições que obtemos são conhecidas como “condições de Kuhn-Tucker para um máximo local”.

Utilizaremos a seguinte representação condensada para as matrizes acima:

$$\begin{pmatrix} -\nabla g_{i_1}(x^*) & \nabla h_{j_1}(x^*) & \delta_1 \cdot I \\ -\frac{\partial g_{i_1}}{\partial x_1}(x^*) & \dots & -\frac{\partial g_{i_1}}{\partial x_1}(x^*) & \frac{\partial h_{j_1}}{\partial x_1}(x^*) & \dots & \frac{\partial h_{j_1}}{\partial x_1}(x^*) & \delta_1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial g_{i_l}}{\partial x_n}(x^*) & \dots & -\frac{\partial g_{i_l}}{\partial x_n}(x^*) & \frac{\partial h_{j_l}}{\partial x_n}(x^*) & \dots & \frac{\partial h_{j_l}}{\partial x_n}(x^*) & 0 & \dots & \delta_n \end{pmatrix},$$

$\nabla g_{i_k}(x^*)$ e $\nabla h_{j_k}(x^*)$ representam, nas colunas da matriz, os gradientes das funções que definem as fronteiras ativas e $\delta_k \cdot I$ a matriz quadrada $n \times n$, cuja diagonal principal é definida pelos δ_k ;

$$\begin{pmatrix} \lambda^* \\ -\mu^* \\ -\nu^* \end{pmatrix} = \begin{pmatrix} \lambda_{i_1}^* \\ \vdots \\ \lambda_{i_l}^* \\ -\mu_{j_1}^* \\ \vdots \\ -\mu_{j_l}^* \\ -\nu_1^* \\ \vdots \\ -\nu_n^* \end{pmatrix} \mathbf{e} \left(-\nabla f(x^*) \right) = \begin{pmatrix} -\frac{\partial f}{\partial x_1}(x^*) \\ \vdots \\ -\frac{\partial f}{\partial x_n}(x^*) \end{pmatrix}.$$

Nessa notação temos $\begin{pmatrix} -\nabla g_{i_1}(x^*) & \nabla h_{j_1}(x^*) & \delta_1 \cdot I \end{pmatrix} \cdot \begin{pmatrix} \lambda^* \\ -\mu^* \\ -\nu^* \end{pmatrix} = (-\nabla f(x^*)).$

Utilizando essa notação, redemonstramos o Teorema de Kuhn-Tucker usando uma notação mais condensada

Teorema 2 – Kuhn-Tucker (1950)

i) Se x^* é um máximo local para $f(x)$, sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_1 \geq 0, \dots, x_n \geq 0,$$

tal que x^* satisfaz a qualificação de restrição e $\nabla f(x^*) \neq 0$. Então, x^* satisfaz as condições abaixo, ditas condições de Kuhn-Tucker para um máximo,

$$\begin{aligned} KT1: \quad \nabla f(x^*) &= \sum \lambda_i^* \cdot \nabla g_i(x^*) + \sum \mu_j^* \cdot \nabla h_j(x^*) + \sum \nu_k^* \cdot \bar{e}_k, \\ KT2: \quad \lambda_i^* \cdot (g_i(x^*) - a_i) &= 0, \quad \mu_j^* \cdot (h_j(x^*) - b_j) = 0, \quad \nu_k^* \cdot x_k^* = 0, \\ KT3: \quad \lambda_i^* &\geq 0, \quad \mu_j^* \leq 0, \quad \nu_k^* \leq 0. \end{aligned}$$

ii) Se x^* é um mínimo local para $f(x)$, sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_1 \geq 0, \dots, x_n \geq 0,$$

tal que x^* satisfaz a qualificação de restrição e $\nabla f(x^*) \neq 0$. Então, x^* satisfaz as condições abaixo, ditas condições de Kuhn-Tucker para um mínimo,

$$\begin{aligned} KT1: \quad \nabla f(x^*) &= \sum \lambda_i^* \cdot \nabla g_i(x^*) + \sum \mu_j^* \cdot \nabla h_j(x^*) + \sum \nu_k^* \cdot \bar{e}_k, \\ KT2: \quad \lambda_i^* \cdot (g_i(x^*) - a_i) &= 0, \quad \mu_j^* \cdot (h_j(x^*) - b_j) = 0, \quad \nu_k^* \cdot x_k^* = 0, \\ KT3: \quad \lambda_i^* &\leq 0, \quad \mu_j^* \geq 0, \quad \nu_k^* \geq 0. \end{aligned}$$

Dem:

Digamos que as fronteiras $g_{i_s}(x) = a_{i_s}, \dots, h_{j_s}(x) = b_{j_s}, \dots, x_{n_s} = 0$ sejam ativas e que $x(t)$ seja uma curva admissível em x^* . Nesse caso,

$$(x'(0))' \cdot (-\nabla g_{i_s}(x^*) - \nabla h_{j_s}(x^*) - \delta_k \cdot I) \geq 0,$$

como x^* satisfaz a qualificação de restrição e é um máximo local, então

$(x'(0))' \cdot (-\nabla f(x^*)) \geq 0$. Pelo Lema de Farkas, existe $\begin{pmatrix} \lambda^* \\ -\mu^* \\ -\nu^* \end{pmatrix} \geq 0$, tal que

$$(-\nabla g_{i_s}(x^*) - \nabla h_{j_s}(x^*) - \delta_k \cdot I) \cdot \begin{pmatrix} \lambda^* \\ -\mu^* \\ -\nu^* \end{pmatrix} = -\nabla f(x^*).$$

Da expressão acima, deduz-se as três condições de Kuhn-Tucker para um máximo. A demonstração das condições para mínimo, obtemos observando que se x^* é um mínimo local para $f(x)$, então x^* é um máximo local para $-f(x)$. ■

Usualmente, esse resultado é apresentado através do lagrangeano. Se x^* é um máximo local para $f(x)$, sujeita às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_1 \geq 0, \dots, x_n \geq 0,$$

tal que x^* satisfaz a qualificação de restrição, então

$$KT1: \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = -\sum \lambda_i^* \cdot \frac{\partial g_i}{\partial x_r}(x^*) - \sum \mu_j^* \cdot \frac{\partial h_j}{\partial x_r}(x^*) - \mu_{i+r}^* = 0$$

$$KT2: \lambda_i^* \cdot (g_i(x^*) - a_i) = 0, \mu_j^* \cdot (h_j(x^*) - b_j) = 0, \mu_{i+r}^* \cdot x_r^* = 0,$$

$$KT3: \lambda_i^* \geq 0, \mu_j^* \leq 0.$$

Se omitimos do lagrangeano $-\mu_{i+r} \cdot x_r - \dots - \mu_{i+n} \cdot x_n$ e escrevemos $L(x, \lambda, \mu) = f(x) - \lambda_1 \cdot (g_1(x) - a_1) - \dots - \lambda_k \cdot (g_k(x) - a_k) - \mu_1 \cdot (h_1(x) - b_1) - \dots - \mu_l \cdot (h_l(x) - b_l)$,

então a primeira condição de Kuhn-Tucker pode ser apresentada como:

$$\frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = -\sum \lambda_i^* \cdot \frac{\partial g_i}{\partial x_r}(x^*) - \sum \mu_j^* \cdot \frac{\partial h_j}{\partial x_r}(x^*) \leq 0, \text{ pois } \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = -\sum \lambda_i^* \cdot \frac{\partial g_i}{\partial x_r}(x^*) - \sum \mu_j^* \cdot \frac{\partial h_j}{\partial x_r}(x^*) = \mu_{i+r}^* \leq 0.$$

$$\text{Se } x_r^* > 0, \text{ então, } \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = -\sum \lambda_i^* \cdot \frac{\partial g_i}{\partial x_r}(x^*) - \sum \mu_j^* \cdot \frac{\partial h_j}{\partial x_r}(x^*) = 0.$$

A segunda das condições é apresentada como $x_r^* \cdot \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = 0$, em relação às fronteiras $x_1 \geq 0, \dots, x_n \geq 0$, e a terceira condição permanece inalterada. Em resumo:

$$KT1: \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = -\sum \lambda_i^* \cdot \frac{\partial g_i}{\partial x_r}(x^*) - \sum \mu_j^* \cdot \frac{\partial h_j}{\partial x_r}(x^*) \leq 0$$

$$KT2: \lambda_i^* \cdot (g_i(x^*) - a_i) = 0, \mu_j^* \cdot (h_j(x^*) - b_j) = 0, x_r^* \cdot \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = 0,$$

$$KT3: \lambda_i^* \geq 0, \mu_j^* \leq 0.$$

Para mínimos, expressamos de maneira semelhante, resultando:

$$KT1: \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = -\sum \lambda_i^* \cdot \frac{\partial g_i}{\partial x_r}(x^*) - \sum \mu_j^* \cdot \frac{\partial h_j}{\partial x_r}(x^*) \geq 0$$

$$KT2: \lambda_i^* \cdot (g_i(x^*) - a_i) = 0, \mu_j^* \cdot (h_j(x^*) - b_j) = 0, x_r^* \cdot \frac{\partial L}{\partial x_r}(x^*, \lambda^*, \mu^*) = 0,$$

$$KT3: \lambda_i^* \leq 0, \mu_j^* \geq 0.$$

6.1 Uma versão do lagrangeano

Em Economia, na maioria dos problemas de otimização, as variáveis $x_1, x_2, \dots, x_n$ são tais que $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$. Nesses problemas, eliminamos do lagrangeano os multiplicadores correspondentes às fronteiras $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$.

Considere o problema:

$$\begin{aligned} \max \quad & f(x_1, x_2, \dots, x_n) \\ \text{sa} \quad & g_1(x_1, \dots, x_n) \leq b_1, \dots, g_k(x_1, \dots, x_n) \leq b_k \\ & h_l(x_1, \dots, x_n) \geq c_1, \dots, h_l(x_1, \dots, x_n) \geq c_l \\ & x_1 \geq 0, \dots, x_n \geq 0. \end{aligned}$$

Se definimos o lagrangeano tradicional

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l) = & f(x_1, x_2, \dots, x_n) - \\ & \lambda_1 \cdot (g_1(x_1, x_2, \dots, x_n) - b_1) - \dots - \lambda_k \cdot (g_k(x_1, x_2, \dots, x_n) - b_k) \\ & - \mu_1 \cdot (h_1(x_1, x_2, \dots, x_n) - c_1) - \dots - \mu_l \cdot (h_l(x_1, x_2, \dots, x_n) - c_l), \\ & - \phi_1 \cdot x_1 - \dots - \phi_n \cdot x_n; \end{aligned}$$

as condições de Kuhn-Tucker se expressam como:

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_1}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_1}(x_1, \dots, x_n) - \phi_1 &= 0; \\ \dots \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_n}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_n}(x_1, \dots, x_n) - \phi_n &= 0; \end{aligned}$$

$$\lambda_1 \cdot (g_1(x_1, x_2, \dots, x_n) - b_1) = 0, \dots, \mu_l \cdot (h_l(x_1, x_2, \dots, x_n) - c_l) = 0; \phi_1 \cdot x_1 = 0, \dots, \phi_n \cdot x_n = 0;$$

$$\lambda_1 \geq 0, \dots, \lambda_k \geq 0, \mu_1 \leq 0, \dots, \mu_l \leq 0, \phi_1 \leq 0, \dots, \phi_n \leq 0.$$

Se definimos o lagrangeano como:

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l) = & f(x_1, x_2, \dots, x_n) - \\ & \lambda_1 \cdot (g_1(x_1, x_2, \dots, x_n) - b_1) - \dots - \lambda_k \cdot (g_k(x_1, x_2, \dots, x_n) - b_k) \\ & - \mu_1 \cdot (h_1(x_1, x_2, \dots, x_n) - c_1) - \dots - \mu_l \cdot (h_l(x_1, x_2, \dots, x_n) - c_l), \end{aligned}$$

OU $\mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) = f(\bar{x}) - \lambda_1 \cdot (g_1(\bar{x}) - b_1) - \dots - \mu_l \cdot (h_l(\bar{x}) - c_l).$

A primeira condição de Kuhn-Tucker se expressa como:

$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_1}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_1}(x_1, \dots, x_n) \leq 0; \dots; \frac{\partial \mathcal{L}}{\partial x_n} = \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_n}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_n}(x_1, \dots, x_n) \leq 0;$ em que eliminamos os multiplicadores $\phi_1 \leq 0, \dots, \phi_n \leq 0$.

A segunda condição de Kuhn-Tucker: $\lambda_1 \cdot (g_1(x_1, x_2, \dots, x_n) - b_1) = 0, \dots, \mu_l \cdot (h_l(x_1, x_2, \dots, x_n) - c_l) = 0$, em relação aos multiplicadores $\phi_1 \leq 0, \dots, \phi_n \leq 0$, expressamos a segunda condição como: $\frac{\partial \mathcal{L}}{\partial x_1} \cdot x_1 = \left(\frac{\partial f}{\partial x_1}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_1}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_1}(x_1, \dots, x_n) \right) \cdot x_1 = 0$; ...;

$$\frac{\partial \mathcal{L}}{\partial x_n} \cdot x_n = \left(\frac{\partial f}{\partial x_n}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_n}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_n}(x_1, \dots, x_n) \right) \cdot x_n = 0.$$

A terceira condição se expressa como $\lambda_1 \geq 0, \dots, \lambda_k \geq 0, \mu_1 \leq 0, \dots, \mu_l \leq 0$.

Para o problema de minimização:

$$\begin{aligned} \min \quad & f(x_1, x_2, \dots, x_n) \\ \text{sa} \quad & g_1(x_1, \dots, x_n) \leq b_1, \dots, g_k(x_1, \dots, x_n) \leq b_k \\ & h_1(x_1, \dots, x_n) \geq c_1, \dots, h_l(x_1, \dots, x_n) \geq c_l \\ & x_1 \geq 0, \dots, x_n \geq 0; \end{aligned}$$

A primeira condição de Kuhn-Tucker se expressa como:

$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_1}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_1}(x_1, \dots, x_n) \geq 0; \dots; \frac{\partial \mathcal{L}}{\partial x_n} = \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_n}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_n}(x_1, \dots, x_n) \geq 0;$$

A segunda condição de Kuhn-Tucker se expressa como

$$\lambda_1 \cdot (g_1(x_1, x_2, \dots, x_n) - b_1) = 0, \dots, \mu_l \cdot (h_l(x_1, x_2, \dots, x_n) - c_l) = 0; \frac{\partial \mathcal{L}}{\partial x_1} \cdot x_1 = \left(\frac{\partial f}{\partial x_1}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_1}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_1}(x_1, \dots, x_n) \right) \cdot x_1 = 0; \dots;$$

$$\frac{\partial \mathcal{L}}{\partial x_n} \cdot x_n = \left(\frac{\partial f}{\partial x_n}(x_1, \dots, x_n) - \lambda_1 \cdot \frac{\partial g_1}{\partial x_n}(x_1, \dots, x_n) - \dots - \mu_l \cdot \frac{\partial h_l}{\partial x_n}(x_1, \dots, x_n) \right) \cdot x_n = 0.$$

A terceira condição se expressa como $\lambda_1 \leq 0, \dots, \lambda_k \leq 0, \mu_1 \geq 0, \dots, \mu_l \geq 0$.

Na seção abaixo empregamos as versões acima para expressar as condições de Kuhn-Tucker.

7. O Teorema da Dualidade da Programação Linear

Nesta seção demonstramos que o valor máximo do problema primal é igual ao valor mínimo do problema dual. Esse resultado é conhecido como Teorema da Dualidade e foi primeiramente publicado por David Gale, Harold Kuhn e Albert Tucker no Simpósio Número Zero da Programação Linear¹³. Nesta seção optamos por seguir a demonstração apresentada por Gale no artigo “*The Basic Theorems of Real Linear Equations, Inequalities, Linear Programming and Game Theory*”¹⁴. A razão pela qual apresentamos esse resultado nesta Lição é porque precisamos, para a sua demonstração, do Lema de Farkas.

7.1 Alternativas de Gale

Seja A matriz $m \times n$; por $(A \ I)$ notamos a matriz $m \times (n+m)$, I é a identidade $m \times m$,

$$(A \ I) = \begin{pmatrix} a_{11} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & 0 & \cdots & 1 \end{pmatrix}.$$

Teorema 3 – Gale (1956)

Seja A matriz $m \times n$, $b \neq 0$. O sistema $A \cdot x \leq b$ tem solução $x \geq 0$ se e somente se para todo $y \in \mathbb{R}^m$, $y \geq 0$, tal que $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$.

$\Rightarrow$ Se $y \in \mathbb{R}^m$, $y \geq 0$, tal que $0 \leq y' \cdot A$, então $0 \leq y' \cdot A \cdot x$, $\forall x \geq 0$. Por hipótese existe $x \geq 0$, $A \cdot x \leq b$, então $0 \leq y' \cdot A \cdot x \leq y' \cdot b$, logo $0 \leq y' \cdot b$.

$\Leftarrow$ Digamos que não exista $x \geq 0$, tal que $A \cdot x \leq b$. Então, não existe $z \geq 0$, tal que $(A \ I) \cdot \begin{pmatrix} x \\ z \end{pmatrix} = b$. Por Farkas, existe $y \in \mathbb{R}^m$, tal que $y' \cdot (A \ I) \geq 0$ e $y' \cdot b < 0$, ou seja, $y' \cdot A \geq 0$, $y \geq 0$, $y' \cdot b < 0$, o que é falso por hipótese.

Alternativas de Gale I: Seja A uma matriz $m \times n$, $b \neq 0$, então apenas uma das alternativas abaixo é verdadeira:

- i) Existe $x \geq 0$, $A \cdot x \leq b$;
- ii) Existe $y \geq 0$, tal que $y' \cdot A \geq 0$ e $y' \cdot b < 0$.

Alternativas de Gale II: Seja A uma matriz $m \times n$, $b \neq 0$, então apenas uma das alternativas abaixo é verdadeira:

- i) Existe $x \geq 0$, $A \cdot x \geq b$;
- ii) Existe $y \geq 0$, tal que $y' \cdot A \leq 0$ e $y' \cdot b > 0$.

A Alternativa de Gale I é uma versão do Teorema de Gale facilmente verificável: se existe solução $x \geq 0$, $A \cdot x \leq b$ e $y \geq 0$, tal que $y' \cdot A \geq 0$, então $y' \cdot b \geq 0$. Portanto, não pode existir $y \geq 0$, tal que $y' \cdot A \geq 0$ e $y' \cdot b < 0$. Por outro lado, se existe $y \geq 0$, tal que $y' \cdot A \leq 0$ e $y' \cdot b < 0$, então a proposição $y \geq 0$, tal que $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$ não é válida, logo não pode existir $x \geq 0$, $A \cdot x \leq b$. A Alternativa de Gale II é obtida da Alternativa I, substituindo A por $-A$ na Alternativa I.

Consideremos os sistemas:

$$\text{primal:} \begin{cases} \max & b' \cdot y \\ \text{sa} & A \cdot y \leq a \\ & y \geq 0 \end{cases} \text{ e dual:} \begin{cases} \min & a' \cdot x \\ \text{sa} & A' \cdot x \geq b \\ & x \geq 0. \end{cases}$$

As possibilidades de inexistência de soluções para os problemas acima são as seguintes:

- i) não existe $y \geq 0$, tal que $A \cdot y \leq a$, e não existe $x \geq 0$, tal que $A' \cdot x \geq b$;
- ii) uma das restrições $y \geq 0$, tal que $A \cdot y \leq a$, ou $x \geq 0$, tal que $A' \cdot x \geq b$ tem solução, mas a outra não;
- iii) ambas as restrições $y \geq 0$, tal que $A \cdot y \leq a$ e $x \geq 0$, tal que $A' \cdot x \geq b$, têm solução, mas não existem $\max b' \cdot y$ ou $\min a' \cdot x$.

A opção I é desprovida de sentido: não vamos propor problemas nos quais as restrições sejam vazias!

Exemplo 13:

$$\text{Primal:} \begin{cases} \max & y_1 + y_2 \\ \text{sa} & y_1 + 0 \cdot y_2 \leq 1 \\ & y_1 + 0 \cdot y_2 \leq -1 \\ & y_1 \geq 0, y_2 \geq 0, \end{cases} \text{ e dual:} \begin{cases} \min & x_1 - x_2 \\ \text{sa} & x_1 + x_2 \geq 1 \\ & 0 \cdot x_1 + 0 \cdot x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{cases}$$

Analisemos a segunda situação, pois podemos supor que a restrição de um dos problemas seja vazia, mas a outra não.

Exemplo 14: Seja o problema:

$$\text{primal:} \begin{cases} \max & y_1 + y_2 \\ \text{sa} & y_1 \leq -1 \\ & y_2 \leq -1 \\ & y_1 \geq 0, y_2 \geq 0, \end{cases}$$

não existem pontos $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \geq 0$ satisfazendo a restrição; o problema

$$\text{dual:} \begin{cases} \min & -x_1 - x_2 \\ \text{sa} & x_1 \geq 1 \\ & x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{cases}$$

não é limitado, pois $-x_1 - x_2 \rightarrow -\infty$. Por outro lado, para o problema

$$\text{dual:} \begin{cases} \min & x_1 + x_2 \\ \text{sa} & -x_1 \geq 1 \\ & -x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{cases}$$

não existem pontos $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq 0$ satisfazendo a restrição; o problema

$$\text{primal: } \begin{cases} \max & y_1 + y_2 \\ \text{s.t.} & -y_1 \leq 1 \\ & -y_2 \leq 1 \\ & y_1 \geq 0, y_2 \geq 0, \end{cases}$$

associado não é limitado, pois $y_1 + y_2 \rightarrow +\infty$. Nesse caso, o problema cuja restrição não é vazia, é ilimitado. Essa é uma característica geral desses problemas.

Digamos que **não** exista $y \geq 0$, tal que $A \cdot y \leq a$, então, por **Gale**, existe $x_0 \geq 0$, tal que $x'_0 \cdot A \geq 0$, ou seja, $A' \cdot x_0 \geq 0$, e $x'_0 \cdot a < 0$. Como estamos supondo que existe $x \geq 0$ tal que $A' \cdot x \geq b$, temos que $A' \cdot (x + \lambda \cdot x_0) \geq b$, para $\lambda \geq 0$, e

$$\lim_{\lambda \rightarrow +\infty} a' \cdot (x + \lambda \cdot x_0) = a' \cdot x + \lim_{\lambda \rightarrow +\infty} \lambda \cdot a' \cdot x_0 = -\infty.$$

Portanto, se não existe $y \geq 0$, tal que $A \cdot y \leq a$, e existe $x \geq 0$, tal que $A' \cdot x \geq b$, então para $x \geq 0$, tal que $A' \cdot x \geq b$. Em outras palavras, se não existe $y \geq 0$, tal que $A \cdot y \leq a$, mas existe $x \geq 0$, tal que $A' \cdot x \geq b$, então o segundo problema não é limitado, ou seja, não existe o mínimo.

Por outro lado, se **não** existe $x \geq 0$, tal que $A' \cdot x \geq b$, então, por **Gale**, existe $y \geq 0$, tal que $A \cdot y \leq a$, e $y' \cdot b < 0$. Se existe $y \geq 0$, tal que $A \cdot y \leq a$, temos para $\lambda \geq 0$, e

.

Portanto, se não existe $x \geq 0$, tal que $A' \cdot x \geq b$, então para $y \geq 0$, tal que $A \cdot y \leq a$.

Lema 1: i) Se **não** existe $y \geq 0$, tal que $A \cdot y \leq a$, e existe $x \geq 0$, tal que $A' \cdot x \geq b$, então para $x \geq 0$, tal que $A' \cdot x \geq b$. ii) Se **não** existe $x \geq 0$, tal que $A' \cdot x \geq b$ e existe $y \geq 0$, tal que $A \cdot y \leq a$, então para $y \geq 0$, tal que $A \cdot y \leq a$.

Analisemos a terceira situação. Suponhamos que ambas as restrições não sejam vazias: existem $y \geq 0$ e $x \geq 0$, tal que $A \cdot y \leq a$ e $A' \cdot x \geq b$, então, $y' \cdot A' \cdot x \leq a' \cdot x$ e $y' \cdot b \leq y' \cdot A' \cdot x$, ou seja, $y' \cdot b \leq a' \cdot x$. Nesse caso, ambos os problemas são limitados e podemos encontrar tanto o máximo do problema primal quanto o mínimo do problema dual. Portanto, $\max b' \cdot y = \min a' \cdot x$. Em outras palavras, a terceira situação acima é impossível. Se ambas as restrições são satisfeitas, então existem $\max b' \cdot y$ ou $\min a' \cdot x$.

Uma maneira de verificar que ambos os problemas têm solução é verificar que ambas as restrições são satisfeitas.

Teorema 4

Existem $y \geq 0$ e $x \geq 0$, tal que $A \cdot y \leq a$, e $A' \cdot x \geq b$ se, e só se, $z^* = w^*$.

Dem $\Rightarrow$ suponhamos que $-\infty = \min A' \cdot x$ ou $z^* = -\infty$, então uma das restrições $y \geq 0$, tal que $A \cdot y \leq a$, ou $x \geq 0$, tal que $A' \cdot x \geq b$, não tem solução, pois se ambas fossem satisfeitas teríamos $z^* = w^*$.

$\Leftarrow$ Por outro lado, se $z^* = w^*$, então ambas as restrições são satisfeitas, pois em caso contrário teríamos $-\infty = \min A' \cdot x$ ou $z^* = -\infty$.

Suponhamos que exista $x \geq 0$, tal que $A' \cdot x \geq b$, e que $z^* = w^*$, então existe $y \geq 0$, tal que $A \cdot y \leq a$ e $z^* = w^*$. Nesse caso, como $y' \cdot A' \cdot x \leq a' \cdot x$, então $z^* = w^*$. Portanto, $z^* = w^*$. Em resumo, se o problema

dual:

tem solução, então, o problema

primal:

tem solução.

Da mesma maneira, suponhamos que exista $y \geq 0$, tal que $A \cdot y \leq a$ e que $z^* = w^*$, então existe $x \geq 0$, tal que $A' \cdot x \geq b$. Como $z^* = w^*$, então $z^* = w^*$. Portanto, se o problema

primal:

tem solução, então, o problema abaixo

dual:

tem solução.

Portanto, se verificarmos que um dos problemas tem solução, então o outro também tem.

Teorema 5

O problema primal: tem solução apenas se o problema dual: tem solução.

Se existe $y \geq 0$, tal que $A \cdot y \leq a$, e existe $x \geq 0$, tal que $A' \cdot x \geq b$, então $z^* = w^*$. Se x^* é um ponto ótimo para o problema primal e y^* é um ótimo para o problema

dual, em particular, z^* . Mas, como já intuímos, z^* . Esse é o resultado conhecido como “Teorema da Dualidade” e o demonstraremos a partir das Alternativas de Gale.

Nosso objetivo é mostrar que a inequação $A \cdot y \leq a$ e $A' \cdot x \geq b$, tal que $y \geq 0$ e $x \geq 0$, tal que $A \cdot y \leq a$ e $A' \cdot x \geq b$.

Suponhamos que a desigualdade anterior seja estrita, isto é, $b' \cdot y < a' \cdot x$ para todo $y \geq 0$, tal que $A \cdot y \leq a$, e para todo $x \geq 0$, tal que $A' \cdot x \geq b$. Ora, $b' \cdot y < a' \cdot x$ para todo $y \geq 0$, tal que $A \cdot y \leq a$ e para todo $x \geq 0$, tal que $A' \cdot x \geq b$ se, e só se, não existem $y \geq 0$, $x \geq 0$, tal que $A \cdot y \leq a$, $A' \cdot x \geq b$ e $b' \cdot y < a' \cdot x$. Logo,

não tem solução. Pela **Alternativa de Gale I**, existe $\theta > 0$, tal que

$$b' \cdot y - a' \cdot x < -\theta$$

ou seja, $\theta > 0$, e $z_1' \cdot a - z_2' \cdot b < 0$.

Suponhamos que $\theta = 0$, então $\theta = 0$, e $z_1' \cdot a - z_2' \cdot b = 0$. Ou seja,

$$z_1' \cdot a - z_2' \cdot b = 0$$

Portanto, novamente pela **Alternativa de Gale I**, não existe $\theta > 0$, tal que $\theta > 0$, isto é, não existem $y \geq 0$ e $x \geq 0$, tais que $A \cdot y \leq a$ e $A' \cdot x \geq b$. Por hipótese, essa última afirmação é falsa. Logo, $\theta > 0$.

Desse modo, existem $\theta > 0$, tais que $\theta > 0$, e $z_1' \cdot a - z_2' \cdot b < 0$, ou seja, $\theta > 0$, e $\theta > 0$, o que é absurdo, pois, por hipótese, não existem $y \geq 0$, $x \geq 0$, tal que $A \cdot y \leq a$, $A' \cdot x \geq b$ e $b' \cdot y < a' \cdot x$. Logo,

tem solução, ou seja, existem $x^* \geq 0$, tal que $A' \cdot x^* \geq b$ e $b' \cdot y^* = a' \cdot x^*$, tais que $A \cdot y^* \leq a$. Como $z^* = z_1^* \cdot a - z_2^* \cdot b$, então z^* é o valor máximo do problema primal é o valor mínimo do problema dual.

Teorema 6 – A Dualidade para a Programação Linear

O problema primal: tem solução apenas se o problema dual:
$$\begin{cases} \min & a' \cdot x \in \mathbb{R} \\ \text{s.t.} & A' \cdot x \geq b \\ & x \geq 0 \end{cases}$$
 tem solução. No caso da existência de soluções para ambos, os problemas: .

7.2 A Programação Linear, os lagrangeanos e as selas

Em 1947, num artigo não destinado para publicação, von Neumann¹⁵ propôs que, nos problemas lineares primal e dual associados,

as soluções ótimas x^* e y^* do primeiro e do segundo problema são tais que . Esse resultado passou a ser conhecido como “Teorema da Dualidade”. Von Neumann não obteve uma prova correta do resultado acima porque usou o chamado “Lema de Farkas” erroneamente.¹⁶

Dantzig salientou que no encontro que teve com von Neumann, em 1947, no qual von Neumann lhe expôs o que sabia sobre programação linear e observou que:

Eu não quero que você pense que estou tirando tudo isso da manga no calor do momento, como um mágico. Eu terminei um livro com Oscar Morgenstern em teoria dos jogos. O que estou fazendo é conjecturar que os dois problemas [o primal e o dual] são equivalentes. A teoria que estou delineando para o seu problema é análogo ao que desenvolvemos para jogos.¹⁷

Dantzig ainda acrescenta que “assim, pela primeira vez, soube do Lema de Farkas e acerca da Dualidade”¹⁸.

A demonstração correta da dualidade não apareceu nem como trabalho de von Neumann, nem como trabalho do Dantzig, mas como o artigo “*Linear Programming and the Theory of Games*”¹⁹, de autoria de David Gale (1921-2008), Harold Kuhn e Albert Tucker nos “*Proceedings*” do famoso “Simpósio Zero” sobre programação linear em 1949. Além da dualidade, os autores mostram que os problemas acima podem ser reduzidos ao problema de determinar as estratégias ótimas num determinado jogo.

Consideremos os sistemas primal: e dual: .

Suponhamos que ambos os problemas tenham solução. O lagrangeano associado ao problema primal é , isto é, Observe que , ou seja, os problemas primal e dual têm o mesmo lagrangeano.

As condições de Kuhn-Tucker para o problema primal são:

em que os ,..., servem para lembrar que as restrições ,..., estão implícitas. Observe que os multiplicadores de Lagrange são tais que . A primeira condição se expressa como $A' \cdot x - \mu = b$ ou , $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \leq 0$.

Observe que as equações acima se traduzem como $x' \cdot (A \cdot y - a) = 0$ e . Se é solução do problema primal, então existe um único $x^* \geq 0$, que é o vetor formado com os correspondentes multiplicadores de Lagrange do problema primal que correspondem às inequações $a_{k1} \cdot y_1 + \dots + a_{kn} \cdot y_n \leq a_k$, , e um único vetor $\mu' \leq 0$, que também são multiplicadores e correspondem às inequações . Observe que , ou seja, $x^* \geq 0$ satisfaz as restrições do problema dual. Como $x' \cdot (A \cdot y' - a) = 0$ e $\mu' \cdot y^* = 0$, então . Portanto, $x^* \geq 0$ é uma solução do problema dual.

Por outro lado, como $(x' \cdot A - b') \cdot y^* = \mu' \cdot y^* = 0$, então é o vetor de multiplicadores associados à solução $x^* \geq 0$ do dual.

Enfatizando, dada uma solução $y^* \geq 0$ do problema primal, o vetor $x^* \geq 0$ de multiplicadores de Lagrange associados ao $y^* \geq 0$ é uma solução do problema dual. Por outro lado, $y^* \geq 0$ é o vetor dos multiplicadores de Lagrange associado à solução $x^* \geq 0$ do problema dual.

Teorema 7

Cada solução do problema primal está associado a uma solução $x^* \geq 0$ do problema dual, que é o vetor dos multiplicadores de Lagrange associados à

solução $x^* \geq 0$ do problema primal. Cada solução $x^* \geq 0$ do problema dual está associado a uma solução do problema primal, que é o vetor dos multiplicadores de Lagrange associados à solução $x^* \geq 0$ do problema dual.

Mostremos que as soluções y^* e $x^* \geq 0$, associadas como no teorema acima, constituem uma **sela do lagrangeano** de ambos os problemas.

$\mathcal{L}(y^*, x^*) = b' \cdot y^* + x^{*'} \cdot (a - A \cdot y^*) \leq b' \cdot y^* + x^{*'} \cdot (a - A \cdot y^*) = L(y^*, x^*)$, para $x \geq 0$, pois $x^{*'} \cdot (a - A \cdot y^*) = 0$ e $\mathcal{L}(y^*, x^*) = x^{*'} \cdot a + (b' - x^{*'} \cdot A) \cdot y$, logo, para $y \geq 0$, pois $(x^{*'} \cdot A - b') \cdot y^* = 0$ e $(x^{*'} \cdot A - b') \cdot y \geq 0$. Portanto,

$$\mathcal{L}(y, x^*) \leq \mathcal{L}(y^*, x^*) \leq \mathcal{L}(y^*, x).$$

Finalmente, mostremos que os coeficientes das variáveis de folga no problema primal são valores ótimos para o problema dual e vice-versa. Consideremos a última equação do *tableau* simplex: $F - b' \cdot y = 0$. No *tableau* final, a última equação resulta $F - b' \cdot y^* - x^{*'} \cdot (A \cdot y^* + z^* - a)$, em que $z^* = \begin{pmatrix} z_1^* \\ \vdots \\ z_m^* \end{pmatrix} \geq 0$ é o vetor final das variáveis de folga. Portanto, o vetor ótimo $x^* \geq 0$ é obtido como os coeficientes das variáveis de folga no *tableau* final: $\frac{\partial F}{\partial z_i}(y^*, x^*, z^*) = -x_i^* \leq 0$.

8. O Teorema de Kuhn-Tucker vetorial

Exemplo 15 – A Qualificação de Restrição: Exemplo de C. B. Tompkins²⁰: Um exemplo do artigo de Kuhn e Tucker é o seguinte:

$$\begin{aligned} \max \quad & f_1(x) = x \\ & f_2(x) = 2 \cdot x - x^2 \\ \text{s.a} \quad & \Omega = \{x \in \mathbb{R} \mid h(x) = x \geq 0\}. \end{aligned}$$

Observemos que $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ é um máximo vetorial, pois se $\begin{pmatrix} x \\ 2 \cdot x - x^2 \end{pmatrix} \geq \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} x \\ 2 \cdot x - x^2 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, então, necessariamente, $x > 1$, pois $2 \cdot x - x^2$ tem um máximo em $x = 1$, mas se $x > 1$, então $2 \cdot x - x^2 < 1$. Seja $x: [0, +\infty) \rightarrow \Omega$, tal que, então $\begin{pmatrix} x'(0) \\ 0 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} x'(0) \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ se $x'(0) > 0$.

Se $x_0 > 1$, então $\begin{pmatrix} x_0 \\ 2 \cdot x_0 - x_0^2 \end{pmatrix}$ também é um máximo vetorial, pois se $\begin{pmatrix} x \\ 2 \cdot x - x^2 \end{pmatrix} \geq \begin{pmatrix} x_0 \\ 2 \cdot x_0 - x_0^2 \end{pmatrix}$ e $\begin{pmatrix} x \\ 2 \cdot x - x^2 \end{pmatrix} \neq \begin{pmatrix} x_0 \\ 2 \cdot x_0 - x_0^2 \end{pmatrix}$, então temos duas possibilidades:

$$\begin{aligned}x > x_0 &\rightarrow 2 \cdot x - x^2 < 2 \cdot x_0 - x_0^2; \\2 \cdot x - x^2 > 2 \cdot x_0 - x_0^2 &\rightarrow x_0 < x.\end{aligned}$$

Seja $x: [0, +\infty) \rightarrow \Omega$, tal que $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix} = \begin{pmatrix} x'(0) \\ (2-2 \cdot x_0) \cdot x'(0) \end{pmatrix}$. Se $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, se $(2-2 \cdot x_0) \cdot x'(0) > 0$, então $x'(0) < 0$.

Portanto, existe uma diferença entre o caso $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} x_0 \\ 2 \cdot x_0 - x_0^2 \end{pmatrix}$, $x_0 > 1$. No primeiro caso, podemos aumentar $(f_1 \circ x)'(0) = x'(0)$ sem diminuir $(f_2 \circ x)'(0) = 0$. No segundo caso, $x_0 > 1$, se aumentar $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix}$, então diminui $(f_2 \circ x)'(0)$ e vice-versa. Kuhn e Tucker conceituam essas duas situações chamando $\begin{pmatrix} x_0 \\ 2 \cdot x_0 - x_0^2 \end{pmatrix}$, $x_0 > 1$, um máximo vetorial próprio, pois não existe $x: [0, +\infty) \rightarrow \Omega$, curva admissível, tal que $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Em outras palavras, em relação às derivadas $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix}$ das curvas admissíveis, $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ é máximo vetorial.

Nesse sentido, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ não é próprio, pois se $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Ou seja, $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ não é um máximo vetorial em relação às derivadas $\begin{pmatrix} (f_1 \circ x)'(0) \\ (f_2 \circ x)'(0) \end{pmatrix}$. ■ Kuhn e Tucker limitam a extensão do teorema anterior aos máximos vetoriais próprios. Seja $x: [0, +\infty) \rightarrow \Omega$ uma trajetória admissível em x^* satisfazendo às restrições de qualificação. Para todo $t > 0$, necessariamente existe j , $1 \leq j \leq n$, tal que $f_j(x(t)) \leq f_j(x^*)$. Isso implica que existe $\bar{t}$, $0 < \bar{t} < t$, tal que $\left. \frac{d}{dt} f_j(x(t)) \right|_{t=\bar{t}} = \nabla f_j(x(\bar{t})) \cdot x'(\bar{t}) \leq 0$. Fazendo $\bar{t} \rightarrow 0^+$, concluímos que existe j , $1 \leq j \leq n$,

$$\left. \frac{d}{dt} f_j(x(t)) \right|_{t=0} = \nabla f_j(x^*) \cdot x'(0) \leq 0.$$

Se, entretanto, existe $\nabla f_i(x^*) \cdot x'(0) > 0$, então existe $\bar{t}$, pela condição de ser x^* um máximo vetorial próprio.

Digamos que as fronteiras $g_i(x) = a_i, \dots, g_i(x) = a_i, h_i(x) = b_i, \dots, h_i(x) = b_i, \dots, x_{n_1} = 0, \dots, x_{n_p} = 0$ sejam ativas. E ainda, suponhamos que $\left. \frac{d}{dt} f_{p_1}(x(t)) \right|_{t=0} = \nabla f_{p_1}(x^*) \cdot x'(0) > 0, \dots, \left. \frac{d}{dt} f_{p_k}(x(t)) \right|_{t=0} = \nabla f_{p_k}(x^*) \cdot x'(0) > 0$. Notemos por $\nabla f_{p_k}(x^*)$, os gradientes dessas funções f_{p_k} , tais que $\nabla f_k(x^*) \cdot x'(0) \leq 0$ para todo k . Esse conjunto pode ser vazio, ou seja, $\nabla f_k(x^*) \cdot x'(0) \leq 0$ para todo k . As

derivadas restantes são $\nabla f_j(x^*) \cdot x'(0) \leq 0$. Digamos que $\left. \frac{d}{dt} f_{q_1}(x(t)) \right|_{t=0} = \nabla f_{q_1}(x^*) \cdot x'(0) \leq 0, \dots$. Notemos por $\nabla f_{p_a}(x^*)$, os gradientes dessas funções, tais que. Portanto,

$$(x'(0))' \cdot (-\nabla g_k(x^*) - \nabla h_{j_a}(x^*) - \delta_k \cdot I - \nabla f_{p_k}(x^*)) \geq 0.$$

Como x^* é um máximo vetorial, sendo assim existem derivadas tais $\left. \frac{d}{dt} f_j(x(t)) \right|_{t=0} = \nabla f_j(x^*) \cdot x'(0) \leq 0$. Pelo Lema de Farkas, existe

,

tal que

A expressão acima está afirmando que

$$\begin{aligned} \nabla f_{d_1}(x^*) &= \sum \lambda_{d_1}^* \cdot \nabla g_l(x^*) + \sum \mu_{j_{d_1}}^* \cdot \nabla h_j(x^*) + \sum \nu_{k_{d_1}}^* \cdot \bar{e}_k + \sum \theta_{d_1} \cdot \nabla f_l(x^*); \\ &\vdots \\ \nabla f_{d_n}(x^*) &= \sum \lambda_{d_n}^* \cdot \nabla g_l(x^*) + \sum \mu_{j_{d_n}}^* \cdot \nabla h_j(x^*) + \sum \nu_{k_{d_n}}^* \cdot \bar{e}_k + \sum \theta_{d_n} \cdot \nabla f_l(x^*). \end{aligned}$$

Somando as expressões acima obtemos:

Ou ainda,

8.1 O Teorema de Kuhn-Tucker para a otimização vetorial

Relembremos as definições de mínimo e máximo vetorial dadas na Lição Um. Seja um conjunto de funções $f_1, \dots, f_p: \Omega \rightarrow \mathbb{R}$. Dizemos que $x^* \in \Omega$ é um **mínimo vetorial** se não existe x viável, isto é, $x \in \Omega$, tal que

e ;

$x^* \in \Omega$ é um **máximo vetorial** $x^* \in \Omega$ se não existe x viável, isto é, $x \in \Omega$, tal que

$$\begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \leq \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_p(x^*) \end{pmatrix} \text{ e } \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \neq \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_p(x^*) \end{pmatrix};$$

Em outras palavras, **se não é possível diminuir alguns dos valores sem que isso implique aumento em alguns dos demais**, chamamos x^* um **mínimo vetorial**. De forma similar, se não é possível aumentar alguns dos valores sem que isso implique diminuição em outros valores, dizemos que x^* é um **máximo vetorial**.

Seja um conjunto de funções $f_1, \dots, f_p$. Dizemos que $x^* \in \Omega$ é um **mínimo vetorial fraco** se não existe x viável, isto é, $x \in \Omega$, tal que $f_i(x) < f_i(x^*)$ para todo $i = 1, \dots, p$. Em outras palavras, **se não é possível diminuir todos os valores**.

Dizemos que x^* um **máximo vetorial fraco** se não existe x viável, isto é, $x \in \Omega$, tal que $f_i(x) > f_i(x^*)$ para todo $i = 1, \dots, p$, ou seja, **se não é possível aumentar todos os valores**.

Se não é possível diminuir todos os valores, então não é possível diminuir alguns sem aumentar outros. Portanto, todo o mínimo vetorial é um mínimo vetorial fraco. Isso justifica o adjetivo “fraco”, pois o conceito amplia o conjunto dos mínimos acrescentando mínimos que não são “fortes”, isto é, mínimos vetoriais conforme a definição da Lição Um. O conjunto dos mínimos vetoriais está contido no conjunto dos mínimos vetoriais fracos. O mesmo é válido para máximos vetoriais e máximos vetoriais fracos. Sem dúvida, um máximo vetorial ou um mínimo vetorial são também máximo ou mínimo fracos. Por exemplo, **se não é possível aumentar digamos f_1 sem que algum $f_j(x) < f_j(x^*)$, então não é possível aumentar todos os f_i simultaneamente.** Portanto, todo o máximo ou mínimo vetorial é máximo ou mínimo fraco.

A proposição oposta não é verdadeira em geral, pois pode ocorrer que não possamos aumentar **todas**, porém podemos aumentar algumas e, nesse caso, o máximo fraco não é um máximo vetorial, ou seja, eficiente no sentido de Pareto. Por exemplo, no conjunto, $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ é um máximo fraco, mas não é um máximo vetorial. Igualmente, um mínimo fraco pode não ser um mínimo: $\Omega = \left\{ \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1/2 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 1/2 \end{smallmatrix}\right) \right\}$, $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ é um mínimo fraco.

Como anteriormente, Ω é uma região definida por $g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l$. Observemos que se x^* é um máximo vetorial, então para qualquer $x \in \Omega$ ou, para todo $i, 1 \leq i \leq n$, ou caso exista $i, 1 \leq i \leq n$, tal que, então existe $j, 1 \leq j \leq n$, tal que $f_j(x) < f_j(x^*)$.

Teorema 8 – Kuhn-Tucker Vetorial (1950)

i) Se x^* é um máximo vetorial próprio para $f_1, \dots, f_p: \Omega \rightarrow \mathbb{R}$ sujeitas às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, x_1 \geq 0, \dots, x_n \geq 0.$$

Então, x^* satisfaz as condições abaixo, ditas condições de Kuhn-Tucker para um máximo vetorial,

ii) Se x^* é um mínimo vetorial próprio para $f_1, \dots$, sujeitas às restrições

$$g_1(x) \leq a_1, \dots, g_k(x) \leq a_k, h_1(x) \geq b_1, \dots, h_l(x) \geq b_l, \dots, x_n \geq 0,$$

então x^* satisfaz as condições abaixo, ditas condições de Kuhn-Tucker para um mínimo,

Uma $f(x)$ é dita **monótona** se

$$e, \text{ então } f \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} < f \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Ou seja, f é tal que basta uma componente do vetor y ser maior que a correspondente componente do vetor x para que $f(y) > f(x)$.

Por exemplo, funções de utilidade, como a velha conhecida Cobb-Douglas ou , possuem essa propriedade. Se algum dos bens tem a quantidade aumentada, aumenta a utilidade.

Observe que a propriedade da monotonicidade não é satisfeita, por exemplo, por utilidades de Leontief, quando os bens são complementares perfeitos: aumentar a quantidade de um bem apenas, não implica aumento de utilidade.

Exemplo 16: Seja a função , cujas algumas isoquantas estão no desenho abaixo:

Os vértices das isoquantas estão sobre a reta $x_2 = \frac{3}{2} \cdot x_1$. A isoquanta está destacada no gráfico e o seu vértice é o ponto $(3, \frac{3}{2})$. Utilidades da forma , que representam pares de bens complementares, não são monótonas, pois o aumento da quantidade de apenas um dos bens não representa aumento de utilidade.

Seja um conjunto de funções , **contínuas** e **monótonas**. Uma **alocação viável** é um vetor , em que cada $x_i \in \Omega$. A interpretação é simples: cada um dos agentes usufrui uma utilidade $f_i(x_i), i=1, \dots, m$ ao consumir a cesta $x_i \in \Omega$, . Nessa situação em que as utilidades são contínuas e monótonas, toda a alocação que é um máximo fraco é também um máximo vetorial. Suponhamos que é um máximo fraco e não é **um máximo vetorial**, então existe uma alocação viável, isto é, $x_i \in \Omega$, tal que

$$e \begin{pmatrix} f_1(x_1) \\ \vdots \\ f_m(x_m) \end{pmatrix} \neq \begin{pmatrix} f_1(x_1^*) \\ \vdots \\ f_m(x_m^*) \end{pmatrix}.$$

Como estamos supondo que $x^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \end{pmatrix}$ **não é** um máximo vetorial, então existe $\theta > 0$ e para as demais componentes temos $x_i^* > 0$, ou seja, é possível aumentar uma das componentes sem diminuir as demais. Como f_i é contínua, então podemos escolher δ , tal que $f_i(x_i - \theta \cdot x_i^*) > f_i(x_i^*)$, pois $x_i^* > 0$. Observe que estamos retirando uma pequena quantidade de cada bem na cesta do consumidor mantendo a utilidade da cesta maior que a utilidade da cesta x^* . Pode ocorrer que $x_j = 0$, mas não representa problema, pois $f_j(x_j) = 0$. Como as funções são monótonas, então para cada uma das demais funções f_j . Dividimos a cesta retirada do consumidor entre os demais consumidores, porque sempre existem restrições nas quantidades disponíveis dos bens. Logo, foi possível aumentar todas as funções simultaneamente e, portanto, **não pode ser** um máximo fraco. Desse modo, nessa situação particular, se x^* é um máximo fraco, então x^* é um máximo vetorial. Em outras palavras, os dois conceitos geram os mesmos vetores.

Observe que, se as utilidades são tais que alguns bens são complementares para alguns consumidores, então a argumentação acima não é válida, pois os acréscimos de algumas quantidades positivas de alguns bens nas cestas desses consumidores não necessariamente implicam aumento de utilidade. Pode estar ocorrendo que nas cestas de alguns consumidores tenham sido acrescidas quantidades positivas de certos bens, mas nenhuma dos bens complementares e, nesses casos, as utilidades não aumentam.

No entanto, se restringirmos as cestas viáveis ao $\Omega = \{x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^n \mid x_j > 0\} = \mathbb{R}_{++}^n$, não precisamos exigir que as utilidades sejam monótonas no sentido restritivo acima, pois as cestas viáveis são tais que os consumidores consomem quantidades positivas de todos os bens e podemos definir o conceito de função monótona de maneira mais restritiva:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} < \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ então } f\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) < f\left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\right),$$

isto é, se as quantidades de todos os bens são aumentadas, então a utilidade aumenta. As funções monótonas anteriores satisfazem essa propriedade, obviamente. Mas, essa última propriedade também é válida para utilidades de Leontief. E, na demonstração anterior, é um vetor que tem quantidades positivas em relação a todos os bens.

Exercício 11

Escreva as condições de Kuhn-Tucker para o problema das alocações eficientes no sentido de Pareto do problema:

no caso em que os dois agentes recebam quantidades **positivas**, de ambos os bens.

i) Verifique que as soluções satisfazem $\frac{\frac{\partial u}{\partial x}(x_1^*, y_1^*)}{\frac{\partial u}{\partial y}(x_1^*, y_1^*)} = \frac{\frac{\partial v}{\partial x}(x_2^*, y_2^*)}{\frac{\partial v}{\partial y}(x_2^*, y_2^*)}$, o que corresponde à curva de contrato na Caixa de Edgeworth;

ii) Fazendo $\theta_1 = \frac{\partial u}{\partial x}(x_1^*, y_1^*)$, e $\lambda_2 = \frac{\partial u}{\partial y}(x_1^*, y_1^*) \cdot \frac{\partial v}{\partial x}(x_2^*, y_2^*)$, observe que todo o ponto na curva de contrato é um ótimo de Pareto.

Exercício 12

Escreva as condições de Kuhn-Tucker para o problema das alocações eficientes no sentido fraco de Pareto, dadas pelo problema:

no caso em que os dois agentes recebam quantidades **positivas**, de ambos os bens.

i) Comprove que as soluções do problema acima correspondem a máximos vetoriais fracos;

ii) Verifique que as soluções satisfazem $\frac{\partial u}{\partial x}(x_1^*, y_1^*) = \frac{\partial v}{\partial x}(x_2^*, y_2^*)$. Por que, indiretamente, isso comprova que, nesse caso particular, os máximos fracos são máximos fortes?

Exercício 13

A expressão $\nabla f(x(t)) \cdot x'(t)$ é chamada derivada direcional de f em $x(t)$ na direção do vetor $x'(t)$. Não é uma expressão muito adequada, pois depende de $x(t)$ e não apenas do ângulo $\theta(t)$ entre $\nabla f(x(t))$ e $x'(t)$. Por essa razão, alguns autores preferem definir a derivada direcional apenas para cursores, ou seja, vetores de norma 1. Se u é tal que $|u|=1$ e $x(t)$ é uma curva tal que $x'(t)=u$, a derivada direcional de f em $x(t)$ na direção do vetor u é

,

em que θ é o ângulo entre ∇f e u . Nós não utilizaremos essa definição, mas a anterior, pois estamos interessados apenas se o número $\nabla f(x(t)) \cdot u$ é negativo, zero ou positivo e não no seu valor absoluto. Usando derivadas direcionais responda:

i) Se a trajetória $x(t)$ é tal que $\vec{x}'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))$ são ortogonais, então está contida numa isoquanta de f ? Por quê?

ii) se a trajetória $x(t)$ é tal que $\vec{x}'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))$ e formam um ângulo $0 < \theta < \frac{\pi}{2}$, então se $t_1 < t_2$? Por quê?

iii) se a trajetória $x(t)$ é tal que $\vec{x}'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))$ e formam um ângulo $\theta > \frac{\pi}{2}$, então se $t_1 < t_2$? Por quê?

Exercício 14

Usando derivadas direcionais, justifique porque x^* é um ponto de mínimo ou máximo local do problema

$$\min/\max f(x_1, x_2, \dots, x_n) \quad \text{S.a.} \quad g(x_1, \dots, x_n) \leq a$$

e , então .

Exercício 15

Considerando o problema de minimização de custos sujeito às restrições $f(x_1, x_2) \geq y$, $x_1 > 0$, $x_2 > 0$, . Dado o correspondente lagrangeano $\mathcal{L} = w_1 \cdot x_1 + w_2 \cdot x_2 - \lambda \cdot (f(x_1, x_2) - y) - \lambda_1 \cdot x_1 - \lambda_2 \cdot x_2$, utilizando as condições de Kuhn-Tucker, mostre que o ponto ótimo (x_1^*, x_2^*) satisfaz as condições:

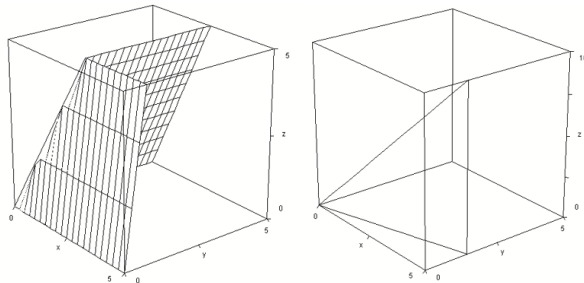
,

,

,

justificando sua resposta.

Exercício 16



i) O gráfico acima é de . Funções do tipo , são ditas de Leontief. Procure entender o gráfico acima. Considere o problema de maximização da utilidade

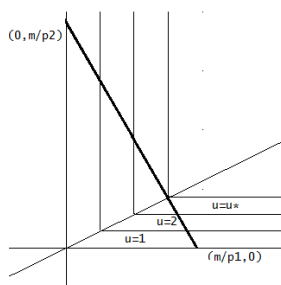
$$\max u(x_1, x_2)$$

s.a.

,

$$x_1 \geq 0, x_2 \geq 0.$$

Obtenha as demandas marshallianas e pelos bens 1 e 2. Observe que nesse caso não se aplica o lagrangeano e, conseqüentemente, as condições de Kuhn-Tucker para um máximo global, pois a função $u(x_1, x_2)$ não é diferenciável. Abaixo, temos um gráfico com algumas isoquantas de e a restrição orçamentária, $x_1 \geq 0, x_2 \geq 0$.



A utilidade é dada por , tal que $-2 < -\frac{a}{b} < -1$.

i) Determine as demandas se e se . Observe que a solução obtida satisfaz as condições de Kuhn-Tucker para um máximo global. Esse é um problema de programação linear.

ii) Se a utilidade é . Esboce os gráficos das isoquantas (que estão no desenho abaixo!) de e verifique que essas cortam o eixo x_1 na cesta $(\sqrt{2 \cdot c}, 0)$, com taxa marginal de substituição ; abaixo, temos um gráfico de com uma isoquanta em destaque no gráfico da direita:

Abaixo, um gráfico mostrando a isoquanta :

Determine a cesta ótima se $p_2 = p_1$; determine a cesta ótima se $p_2 = 3 \cdot p_1$; digamos que , mostre que a cesta ótima é

$$e \quad x_2^* = \frac{\bar{p}_1 - p_2}{p_2} \cdot \frac{m}{2 \cdot \bar{p}_1 - p_2}.$$

Observe que as soluções obtidas satisfazem as condições de Kuhn-Tucker para um máximo global.

9. Tucker e Kuhn

Albert W. Tucker (1905-1995) nasceu no Canadá. Em 1932, doutorou-se em Topologia, uma especialização matemática, com Solomon Lefschetz (1884-1972), um célebre matemático da Universidade de Princeton de origem russa. A vida acadêmica de Tucker esteve sempre ligada a Princeton, onde foi professor desde 1933 até a sua aposentadoria em 1974. Tucker teve três alunos notáveis: Harold William Kuhn (1925-2014), David Gale (1921-2008) e John Forbes Nash (1928-2015).

Os trabalhos de Tucker em Matemática Aplicada devem-se a um acontecimento fortuito. Quando Dantzig esteve em Princeton para encontrar-se com von Neumann, em 1948, este relatou-lhe seu trabalho em Programação Linear para a Força Aérea. Kuhn reconheceu que o trabalho de Dantzig tinha alguma relação com as chamadas “Leis de Kirchhoff” da Eletricidade. Pouco depois, Tucker foi convidado para participar de um projeto de pesquisa para analisar as relações entre a programação linear e os jogos matriciais, financiado pelo *Office Naval Research (ONR)*. Tucker convidou Kuhn e Gale. A pesquisa dos três resultou no Teorema da Dualidade, publicado no artigo “*Linear Programming and the Theory of Games*” referido anteriormente.

Depois desse trabalho inicial, Tucker resolveu explorar a dualidade em programação quadrática – Gale não quis participar – e o trabalho se concretizou no artigo de 1950, “*Nonlinear Programming*”, de autoria de Kuhn e Tucker. Nesse artigo, pela primeira vez aparece o nome “*Nonlinear Programming*”.²¹

Segundo Tinne Kjeldsen²², Dantzig no encontro com von Neumann tinha um outro objetivo além de discutir a Programação Linear, que era a analisar

a viabilidade de o ONR financiar pesquisas nas universidades. A relação de Tucker com o ONR se prolongou de 1948 até 1972.

Tucker é responsável pelo célebre “Dilema do Prisioneiro” da Teoria dos Jogos.

Harold William Kuhn (1925-2014) nasceu no EUA. Um trabalho bem conhecido de Kuhn é o chamado “Método Húngaro”²³ para encontrar soluções para um problema de transporte particular chamado “o problema da designação”.

¹ KUHN, H.; TUCKER, A. “Nonlinear Programming”. In: NEYMAN, Jerzy (Ed.). **Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1950, pp. 481-492.

² KUHN, H. “Nonlinear Programming: a Historical View”. In: LEMKE, C. E.; COTTLE, R. W. **Nonlinear Programming**. New York: AMS, v. 9 (1976), p. 2.

³ KUHN, H. “Nonlinear Programming: A Historical View”. In: COTTLE, Richard W.; LEMKE, C. E. (Ed.). **Nonlinear Programming: Proceedings of a Symposium in Applied Mathematics Held in New York City, SIAM-AMS Proceedings**, v. 9. Providence: American Mathematical Society, 1976, pp. 1-26.

KJELDSEN, T. H. “A Contextualized Historical Analysis of the Kuhn-Tucker Theorem in Nonlinear Programming: the Impact of World War II”. **Historia Mathematica**, v. 27 (2000), pp. 331-361.

⁴ KUHN, H. “Nonlinear Programming: A Historical View”. In: COTTLE, Richard W.; LEMKE, C. E. (Ed.). **Nonlinear Programming: Proceedings of a Symposium in Applied Mathematics Held in New York City, SIAM-AMS Proceedings**, v. 9. Providence: American Mathematical Society, 1976, pp. 1-26.

⁵ KUHN, H; TUCKER, A. “Nonlinear Programming”. In: NEYMAN, Jerzy (Ed.). **Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1950, p. 483.

⁶ PREKOPA, A. “On the Development of Optimization Theory”. **The American Mathematical Monthly**, v. 87, nº 7 (1980), pp. 528, 531.

⁷ KUHN, H. “Nonlinear Programming: a Historical View”. In: LEMKE, C. E.; COTTLE, R. W. **Nonlinear Programming**. New York: AMS, v. 9, 1976, p. 14.

⁸ Ibidem, p. 2.

- ⁹ SLATER, M. "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming". **Cowles Foundation Discussion Paper**, n° 80 (1950).
- ¹⁰ ARROW, K.; HURWICZ, L.; UZAWA, H. "Constraint Qualifications in Maximization Problems". **Naval Research Logistics**, v. 8, n° 2 (1961).
- ¹¹ ARROW, K.; UZAWA, H. **Constraint Qualifications in Maximization Problems II**. Technical Report n. 84, Institute for Mathematical Studies in Social Sciences, Stanford University, 1960.
- ¹² PETERSON, D. "A Review of Constraint Qualifications in Finite-Dimensional Spaces". **SIAM Review**, v. 15, n° 3 (1973), pp. 639-654.
- ¹³ GALE, D.; KUHN, H.; TUCKER, A. "Linear Programming and Theory of Games". In: KOOPMANS, T. **Activities Analysis of Production and Allocation**. New York: John Wiley & Sons, 1951, pp. 317-329.
- ¹⁴ GALE, D. "The Basic Theorems of Real Linear Equations, Inequalities, Linear Programming, and Game Theory". **Naval Research Logistics Quarterly**, v. 3, n° 3 (1956), pp. 193-200.
- ¹⁵ NEUMANN, J. "Discussion of a Maximum Problem". In: TAUB, A. H. (Ed.). **John von Neumann, Collected Works**, v. VI. New York: Pergamon Press, 1963, pp. 89-95.
- ¹⁶ KJELDSSEN, T. H. "A Contextualized Historical Analysis of the Kuhn-Tucker Theorem in Nonlinear Programming: the Impact of World War II". **Historia Mathematica**, v. 27 (2000), pp. 354-355.
- ¹⁷ DANTZIG, G. "Reminiscences About the Origins of Linear Programming". **Operations Research Letters**, v. 1, n° 2 (1982), p. 45.
- ¹⁸ DANTZIG, G. "Reminiscences About the Origins of Linear Programming". **Operations Research Letters**, v. 1, n° 2 (1982), p. 45.
- ¹⁹ GALE, D.; KUHN, H.; TUCKER, A. "Linear Programming and the Theory of Games". In: KOOPMANS, T. C. (Ed.), **Activity Analysis of Production and Allocation**. New York: John Wiley and Sons, INC., 1951, pp. 317-328.
- ²⁰ KUHN, H.; TUCKER, A. "Nonlinear Programming". In: NEYMAN, Jerzy (Ed.). **Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1950, p. 488.
- ²¹ KUHN, H.; TUCKER, A. "Nonlinear Programming". In: NEYMAN, Jerzy (Ed.). **Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1950, pp. 481-492.
- ²² KJELDSSEN, T. H. "The Emergence of Nonlinear Programming: Interactions between Practical Mathematics and Mathematic Proper". **Mathematical Intelligencer**, v. 22, n° 3 (2000), pp. 50-54.
- ²³ KUHN, H. W. "The Hungarian Method for the Assignment Problem". **Naval Research Logistics Quarterly**, v. 2, n° 1-2, pp. 83-97.

Lição SEIS

Convexidade

Sumário

Lição 6 – Convexidade

1. Introdução
2. Subconjuntos convexos do
3. Funções convexas e côncavas
4. Pontos críticos de funções convexas e côncavas
5. Funções quase-côncavas e funções quase-convexas.
6. Teorema de Kuhn-Tucker restrito a funções quase-côncavas e quase-convexas
7. O resultado principal do artigo “Nonlinear Programming” (1950)
8. As qualificações de restrições

Lição 6

Convexidade

1. Introdução

O objetivo desta lição é definir funções convexas, côncavas, quase-convexas e quase-côncavas e obter condições suficientes para mínimo e máximo globais no Teorema de Kuhn-Tucker, quando as funções envolvidas satisfazem determinadas condições de convexidade.

No fim da lição, exporemos o conteúdo do artigo “*Nonlinear Programming*”, de Kuhn e Tucker. A questão da qualificação de restrições.

2. Subconjuntos convexos do $\mathbb{R}^n$

Um subconjunto $S \subseteq \mathbb{R}^n$ é dito um conjunto convexo se dados dois pontos quaisquer pertencentes a S então os pontos entre estes pontos também pertencem a S . Mais formalmente, um subconjunto $S \subseteq \mathbb{R}^n$ é dito **convexo** se

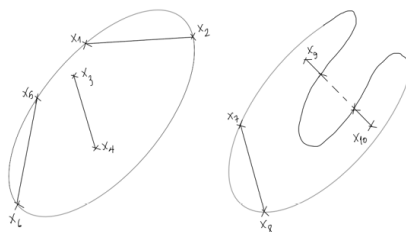
$$(\lambda \cdot x_1 + (1 - \lambda) \cdot x_2) \in S, \forall x_1, x_2 \in S, \lambda \in [0, 1].$$

O segmento de reta constituído pelos pontos entre x_1 e x_2 , dito aberto e notado por (x_1, x_2) , é

$$(x_1, x_2) = \{x \in \mathbb{R}^n \mid x = (\lambda \cdot x_1 + (1 - \lambda) \cdot x_2) \in S, 0 < \lambda < 1\}.$$

Se incluimos x_1 e x_2 , além dos pontos intermediários, temos o segmento fechado $[x_1, x_2]$.

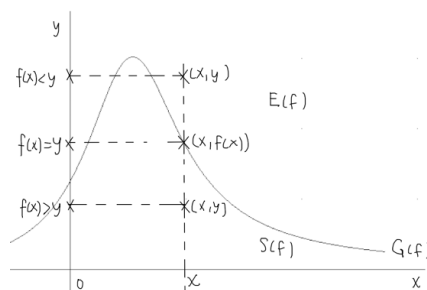
Portanto, s é convexo se dados $\forall x_1, x_2 \in S$, o segmento de reta $[x_1, x_2]$, unindo x_1 e x_2 , está contido em s . No desenho abaixo, o subconjunto da esquerda é convexo, mas o da direita, não:



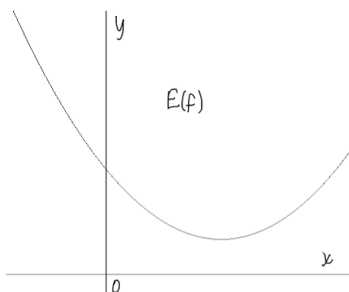
No subconjunto s da esquerda, quaisquer que sejam os pontos $x_i \in S$ e $x_j \in S$, o segmento $[x_i, x_j]$ está sempre contido no subconjunto s . No subconjunto s da direita, existem, por exemplo, $x_9 \in S$ e $x_{10} \in S$, tais que $[x_9, x_{10}] \not\subset S$ não está contido em s . Às vezes dizemos que um conjunto é estritamente convexo se, para dois pontos quaisquer pertencentes a $S \cup S^{front}$, o intervalo (x_1, x_2) está contido no interior do conjunto s .

A teoria moderna da convexidade inicia com os trabalhos dos matemáticos Karl Hermann Brunn (1862-1939) e Hermann Minkowski (1864-1909). A tese de doutorado de Brunn, defendida em 1887, é considerada o primeiro trabalho a tratar de propriedades de conjuntos caracterizados pela convexidade¹. Além da tese de doutorado, Bonnesen e Fenchel destacam um trabalho de Brunn datado de 1889. Minkowski desenvolve a teoria da convexidade num conjunto de trabalhos entre 1897 e 1903. Num artigo muito interessante², Kjeldsen mostra como os dois matemáticos trabalharam o mesmo assunto, porém com diferentes objetivos e metodologias, e como a análise convexa tal como a conhecemos hoje, origina-se das concepções de Minkowski.

O conceito de funções convexas é normalmente atribuído a Johan L. W. V. Jensen (1859-1925), num trabalho de 1905, também num contexto diverso daquele de Brunn e Minkowski^{3 4}. Nas “*Historical Notes*”, no fim da monografia “*Convex Cones, Sets and Functions*”⁵, Fenchel faz uma cronologia detalhada do desenvolvimento da Análise Convexa. Funções convexas e côncavas podem ser definidas inteiramente em termos de conjuntos convexos. Toda a função $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ divide o espaço $\mathbb{R}^{n+1}$, no qual o seu gráfico está mergulhado em três subconjuntos: o gráfico propriamente $G(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid y = f(x)\}$; a parte superior do gráfico $E(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid y \geq f(x)\}$, que chamamos epigráfico de $f(x)$, e a parte inferior do gráfico $S(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid y \leq f(x)\}$, que chamamos de subgráfico de $f(x)$. Veja o desenho a seguir:



Dizemos que uma função $f: S \rightarrow \mathbb{R}$, definida num subconjunto convexo $S \subseteq \mathbb{R}^n$, é **convexa** se $E(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid x \in S, y \geq f(x)\}$ é um subconjunto convexo de $\mathbb{R}^{n+1}$.



Dizemos que uma função $f: S \rightarrow \mathbb{R}$, definida num subconjunto convexo $S \subseteq \mathbb{R}^n$, é **côncava** se $S(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid x \in S, y \leq f(x)\}$ é um subconjunto

convexo $\mathbb{R}^{n+1}$ ou que resulta na mesma, neste caso a função $-f$ é convexa. Se a função f é simultaneamente convexa e côncava, diz-se que f é uma **função afim**. Funções afins são da forma $f(x_1, \dots, x_n) = a_1 \cdot x_1 + \dots + a_n \cdot x_n + b$ e correspondem às funções lineares somadas a uma constante $b \in \mathbb{R}$.

Particularmente para nós da Economia Matemática, a teoria das estruturas convexas teve um grande desenvolvimento a partir da década de 1940. Um poliedro é definido como uma interseção de semiplanos $a_{i1} \cdot x_1 + \dots + a_{in} \cdot x_n + b_i \leq c_i$ para $i=1, \dots, m$, que definem as restrições num problema de minimização ou maximização de uma função afim, como vimos nos problemas de programação linear⁶. A Programação Linear reavivou o interesse no estudo dos poliedros⁷. O Lema de Farkas, que tivemos ocasião de empregar na Lição anterior, pertence à teoria da convexidade.

A teoria da convexidade introduz-se na teoria econômica através do célebre artigo de von Neumann sobre seu modelo de crescimento, publicado na *"The Review of Economic Studies"*, em 1945-1946, com o título *"A Model of General Economic Equilibrium"*. O artigo é anterior a essa data, pois foi publicado na versão original em alemão em 1937, nos anais dos seminários organizados por Karl Menger, em Viena. No título original do artigo, *"Über ein Ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes"*, "Sobre um Sistema de Equações Econômicas e uma Generalização do Teorema do Ponto Fixo de Brouwer" destacava a sua versão do Teorema do Ponto Fixo. Segundo o próprio von Neumann, o trabalho foi apresentado pela primeira vez em 1932, num seminário na Universidade de Princeton⁸. A generalização que von Neumann apresenta é conhecida atualmente como Lema da Interseção de von Neumann. Esse artigo de von Neumann, com muita justiça, é considerado o mais importante artigo de Economia Matemática do século 20. Debru expressa a inflexão que a teoria sofre a partir de

então da seguinte maneira: “[...] a mudança radical nas ferramentas matemáticas na área que estamos tratando se expressa essencialmente pela troca do cálculo pela convexidade e propriedades topológicas [de conjuntos]”⁹.

Os artigos de McKenzie¹⁰ e Debreu e Arrow¹¹, que são os mais célebres trabalhos demonstrando a existência de equilíbrio walrasiano, empregam o conceito de convexidade. A convexidade é essencial para garantir a existência de funções de ofertas das firmas a partir das tecnologias definidas como conjuntos de vetores $Y \subseteq \mathbb{R}^n$ com certas propriedades, dentre elas, a convexidade. Da mesma maneira, na teoria do consumidor, a existência de funções de demanda depende da convexidade das preferências¹².

Na Teoria dos Jogos, a passagem das estratégias puras para as estratégias mistas corresponde à “convexização” dos conjuntos de estratégias puras. Os simplex básicos $\Delta^{n-1} \subseteq \mathbb{R}^n$ são obtidos como a combinação convexa de dos vértices $\bar{0}, e_1, \dots, e_n$, que é o conjunto convexo obtido pela interseção de todos os convexos que contém $\bar{0}, e_1, \dots, e_n$, isto é, o “menor” convexo contendo esses vértices.

Em 1941, o matemático japonês Shizuo Kakutani (1911-2004) publicou uma generalização do Teorema de Brouwer adequada para os objetivos da Teoria Econômica. O Teorema de Kakutani foi empregado em 1950 por John F. Nash Jr.¹³ ¹⁴(1928-2105) para demonstrar a existência, em jogos não cooperativos, dos chamados equilíbrios de Nash. A forma argumentativa empregada por Nash tornou-se padrão na exposição da existência de selas e equilíbrio walrasiano.

O Teorema de Brouwer¹⁵, que continua sendo necessário na argumentação do Kakutani, passou a ser demonstrado através do chamado Teorema KKM (Knaster-Kuratowski-Mazurkiewicz),

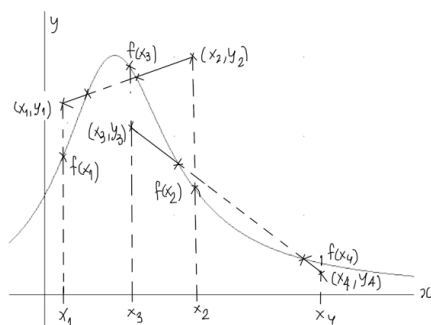
publicado em 1928, que, por sua vez, faz uso do Lema de Sperner (1928), devido ao matemático alemão Emanuel Sperner (1905-1980).

Na teoria da otimização, a convexidade assegura que condições de primeira ordem se tornem condições suficientes, por exemplo, um ponto crítico de uma função convexa diferenciável é um mínimo global. Da mesma maneira, as condições de Kuhn-Tucker para mínimo ou máximo se tornam suficientes com hipóteses adequadas sobre a convexidade das funções envolvidas.

Os conceitos de funções quase-côncavas e quase-convexas foram introduzidos, primeiramente, pelo matemático italiano Bruno de Finetti (1906-1985) em 1949¹⁶. Na seção 3 do artigo de 1928, através da “propriedade K”, von Neumann emprega as definições de funções quase-côncavas e quase-convexas para a demonstração da existência de selas dos *payoffs* esperados. Mas, von Neumann não se refere mais a essas propriedades e nem as destaca como tema de análise por si mesmas¹⁷.

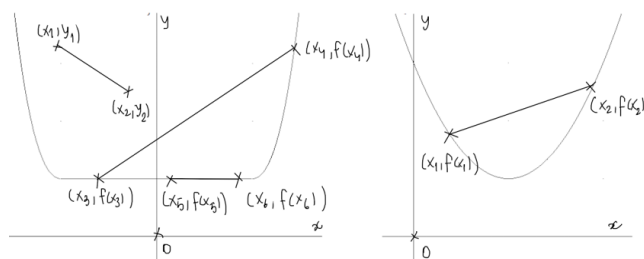
Finalmente, em 1953, Werner Fenchel (1905-1988) introduz o termo “função quase-convexa” nas notas “*Convex Cones, Sets and Functions*” que resultaram dos seminários ministrados em Princeton, em 1953, a convite de Tucker¹⁸.

Olhando o gráfico da função $f(x) = \frac{x+5}{x^2-2 \cdot x+2}$ que está desenhado abaixo, percebemos facilmente que essa função não é nem convexa nem côncava:



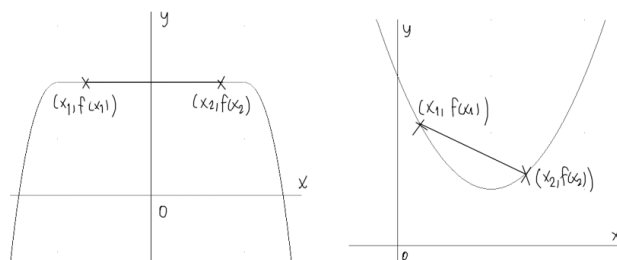
Por exemplo, os pontos $(x_1, y_1) \in E(f)$ e $(x_2, y_2) \in E(f)$, mas o segmento de reta unindo esses pontos não está inteiramente em $E(f)$. Da mesma maneira, os pontos $(x_3, y_3) \in S(f)$ e $(x_4, y_4) \in S(f)$, mas o segmento de reta unindo esses pontos não está inteiramente em $S(f)$.

Os gráficos de funções convexas são gráficos como:



Ambas são convexas porque ambos os epigráficos são subconjuntos convexos do $\mathbb{R}^2$. Todavia, existe uma diferença, a função na esquerda no desenho acima, que tem uma definição complicada $f(x, y) = \left(\frac{\|x\| - 1 + \|x\| - 1}{2 \cdot (\|x\| - 1)} \right) \cdot (1-x)^2 \cdot (x+1)^2 + \frac{1}{2}$, é exemplo de uma função convexa sem ser estritamente convexa, porque os pontos $(x, y) \in E(f)$, entre os pontos $(x_5, f(x_5)) \in E(f)$ e $(x_6, f(x_6)) \in E(f)$, são tais que não estão no interior do subconjunto $E(f)$, mas estão na fronteira desse conjunto. A função na direita, $f(x, y) = (1-x)^2 + \frac{1}{2}$, no desenho acima, quaisquer que sejam os pontos $(x_1, y_1) \in E(f)$ e $(x_2, y_2) \in E(f)$, o segmento unindo estes pontos está sempre no interior do epigráfico, exceto nos extremos (x_1, y_1) e (x_2, y_2) .

As mesmas observações são válidas para funções côncavas se trocarmos os epigráficos pelos subgráficos.



A função da esquerda é côncava sem ser estrita. A função da direita é estritamente côncava.

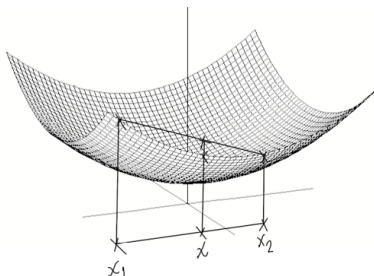
Essas definições gerais servem para estender os resultados sobre a convexidade e a concavidade de funções eliminando propriedades que não são relevantes para as aplicações que se tem em vista em determinado problema. Também possuem um apelo intuitivo que nos permite compreender com facilidade os conceitos de funções convexas e côncavas, mas essa generalidade e apreensão intuitiva cobra seu preço: é muito difícil trabalhar analiticamente com elas. Por isso, faremos uma restrição mais adiante. Nos limitaremos àquelas convexas e côncavas que, em cada ponto do seu gráfico, é possível definir o plano tangente neste ponto.

Mas, antes de fazer isso, apresentaremos versões de definições equivalentes às que acabamos de definir.

3. Funções convexas e côncavas

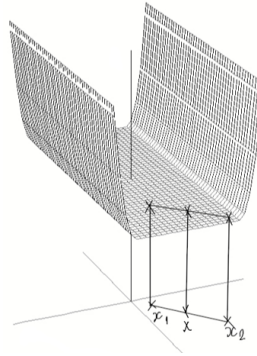
Seja $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, em que $S \subseteq \mathbb{R}^n$ é um subconjunto convexo, diz-se que f é **estritamente convexa** se

$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) < \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2), \forall x_1, x_2 \in S, \lambda \in (0,1).$$



Ou seja, se a reta que une os pontos $(x_1, f(x_1))$ e $(x_2, f(x_2))$ está estritamente sobre a curva $f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2)$, para $\lambda \in (0,1)$, como no desenho acima. No caso do desenho abaixo, a convexidade não é estrita, pois existem pontos

onde a reta está unindo esses pontos, por exemplo, $(x_1, f(x_1))$ e $(x_2, f(x_2))$, situa-se no gráfico da função. Observe que a convexidade estrita corresponde a exigir que o epigráfico da função seja um conjunto estritamente convexo.

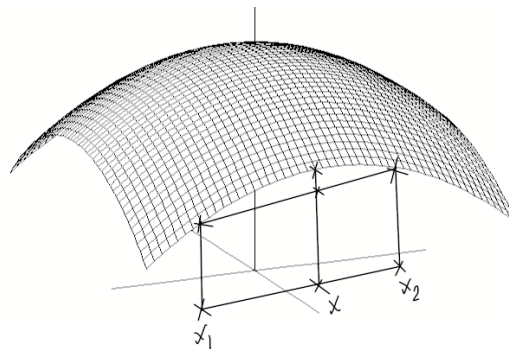


Diz-se que $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é **convexa** se

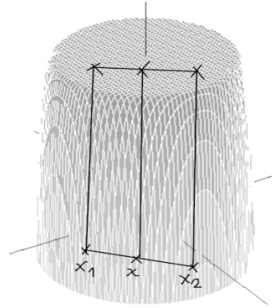
$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2), \forall x_1, x_2 \in S, \lambda \in [0,1].$$

Diz-se que $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é **estritamente côncava** se

$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) < \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2), \forall x_1, x_2 \in S, \lambda \in (0,1).$$



Ou seja, se a reta que une os pontos $(x_1, f(x_1))$ e $(x_2, f(x_2))$ está estritamente sob a curva $f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2)$, para $\lambda \in (0,1)$, como no desenho acima. No caso abaixo, a concavidade não é estrita, pois existem pontos onde a reta está unindo estes pontos, por exemplo, $(x_1, f(x_1))$ e $(x_2, f(x_2))$, está situado no gráfico da função.

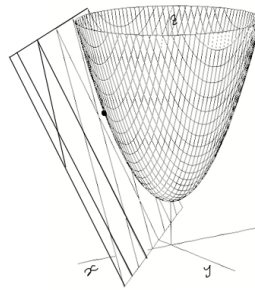


Diz-se que $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é **côncava** se

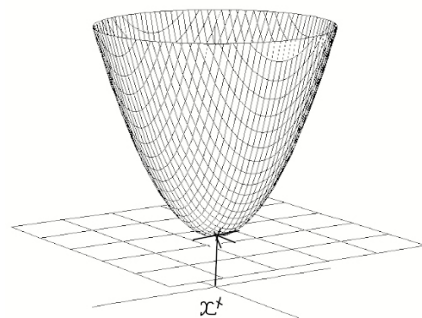
$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2), \forall x_1, x_2 \in S, \lambda \in [0,1].$$

Outra maneira de definir funções convexas e côncavas é através dos planos tangentes. Diz-se que $f: S \rightarrow \mathbb{R}^n$ é **estritamente convexa** se

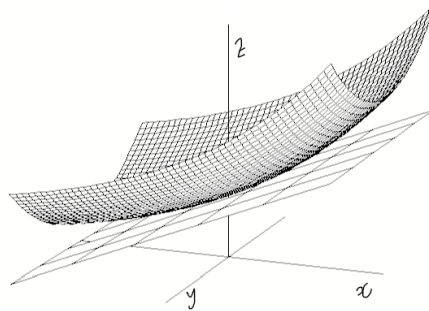
$$f(x) > f(x^*) + \nabla f(x^*) \cdot (x - x^*), x \neq x^*.$$



Em outras palavras, $f: S \rightarrow \mathbb{R}^n$ é estritamente convexa se o gráfico de $f(x)$ está sobre o plano tangente no ponto x^* , obviamente, para pontos $x \neq x^*$. Observe que caso o ponto x^* seja crítico, isto é, $\nabla f(x^*) = 0$, então $f(x) > f(x^*)$. Portanto, x^* é um ponto de mínimo global e único.



Obviamente, uma função estritamente convexa ou convexa pode não ter pontos críticos. Veja o desenho abaixo.



Exemplo 1: $(x-x^*)^2 = (x_1-x_1^*)^2 + (x_2-x_2^*)^2 + \dots + (x_n-x_n^*)^2 > 0$ se $x-x^* \neq 0$, então $x_1^2 + \dots + x_n^2 > (x_1^*)^2 + \dots + (x_n^*)^2 + 2 \cdot x_1^* \cdot (x_1 - x_1^*) + \dots + 2 \cdot x_n^* \cdot (x_n - x_n^*)$. A última desigualdade está assegurando que a função $f(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2$ é estritamente convexa, pois $x_1^2 + \dots + x_n^2 > (x_1^*)^2 + \dots + (x_n^*)^2 + 2 \cdot x_1^* \cdot (x_1 - x_1^*) + \dots + 2 \cdot x_n^* \cdot (x_n - x_n^*)$ está afirmando que $f(x) > f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, para $x - x^* \neq 0$. Obviamente, $f(x_1, \dots, x_n) = c + x_1^2 + \dots + x_n^2$ também é estritamente convexa. $f(x_1, \dots, x_n) = a_1 \cdot x_1^2 + \dots + a_n \cdot x_n^2 + b$ e $a_1 > 0, \dots, a_n > 0$ é estritamente convexa, pois $a_1 \cdot x_1^2 + \dots + a_n \cdot x_n^2 > a_1 \cdot (x_1^*)^2 + \dots + a_n \cdot (x_n^*)^2 + 2 \cdot a_1 \cdot x_1^* \cdot (x_1 - x_1^*) + \dots + 2 \cdot a_n \cdot x_n^* \cdot (x_n - x_n^*)$.

Diz-se que $f: S \rightarrow \mathbb{R}^n$ é **estritamente côncava** se $f(x) < f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, $x \neq x^*$.

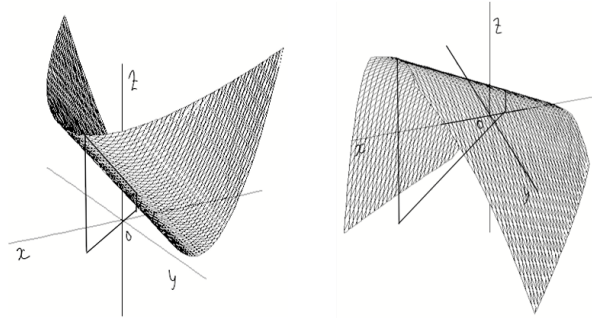
Comentários e desenhos similares aos feitos para funções estritamente convexas são válidos em relação às estritamente côncavas, observando que os gráficos das funções estão sob os planos tangentes e que pontos críticos, caso existam, são mínimos globais únicos para a funções f .

Exemplo 2: Não é difícil perceber que f é estritamente côncava se, e só se, $-f$ é estritamente convexa, pois $f(x) < f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, para $x - x^* \neq 0$, se, e só se, $(-f)(x) > (-f)(x^*) + (-\nabla f)(x^*) \cdot (x - x^*)$, para $x - x^* \neq 0$. Portanto, $f(x_1, \dots, x_n) = -(x_1^2 + \dots + x_n^2)$ é estritamente côncava. $f(x_1, \dots, x_n) = c - x_1^2 - \dots - x_n^2$ também é estritamente côncava. $f(x_1, \dots, x_n) = a_1 \cdot x_1^2 + \dots + a_n \cdot x_n^2 + b$ e $a_1 < 0, \dots, a_n < 0$ é estritamente côncava.

Diz-se que $f: S \rightarrow \mathbb{R}^n$ é **convexa** se $f(x) \geq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, $\forall x$; diz-se que $f: S \rightarrow \mathbb{R}^n$ é **côncava** se $f(x) \leq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, $\forall x$. Abaixo estão os gráficos de uma

convexa e de uma cônica sem serem estritas e sem pontos críticos.

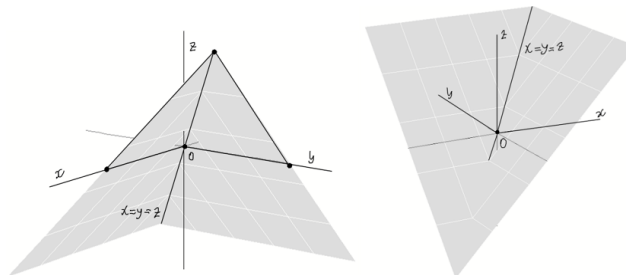
Exemplo 3: $(x-x^*)^2 = (x_1-x_1^*)^2 + (x_2-x_2^*)^2 + \dots + (x_n-x_n^*)^2 \geq 0$, então $x_1^2 + \dots + x_n^2 \geq (x_1^*)^2 + \dots + (x_n^*)^2 + 2 \cdot x_1^* \cdot (x_1 - x_1^*) + \dots + 2 \cdot x_n^* \cdot (x_n - x_n^*)$, para quaisquer $x, x^* \in \mathbb{R}^n$. A última desigualdade está assegurando que a função $f(x_1, \dots, x_n) = a_1 \cdot x_1^2 + \dots + a_n \cdot x_n^2 + b$ é convexa se $a_1 \geq 0, \dots, a_n \geq 0$ é convexa e que $f(x_1, \dots, x_n) = a_1 \cdot x_1^2 + \dots + a_n \cdot x_n^2 + b$ é cônica se $a_1 \leq 0, \dots, a_n \leq 0$. $f(x_1, \dots, x_n) = a_1 \cdot x_1^2 + \dots + a_n \cdot x_n^2 + b$ não é nem convexa, nem cônica se existirem $a_i > 0, a_j < 0$.



Exemplo 4: Funções afins da forma $f(x_1, x_2, \dots, x_n) = a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n + b$ satisfazem ambas as condições para a convexidade e para concavidade, pois, nesses casos, os gráficos das funções são os próprios planos tangentes em cada um dos pontos dos gráficos.

Exercício 1

Verifique que $\min(x, y) = \frac{x+y-|x-y|}{2}$ e que $\max(x, y) = \frac{x+y+|x-y|}{2}$.



$\min(x, y)$ é uma função cônica e $\max(x, y)$ é uma função convexa. Não existem planos tangentes definidos sobre os pontos $x = y = z$, que é uma aresta para ambas as funções. Observe os desenhos acima. Nesses

pontos, as derivadas parciais não estão definidas: $\frac{\min(x+h,x)-\min(x,x)}{h} = \frac{1}{2} - \frac{|h|}{2 \cdot h}$, se $h > 0$, então $\frac{\min(x+h,x)-\min(x,x)}{h} = 0$; se $h < 0$, então $\frac{\min(x+h,x)-\min(x,x)}{h} = 1$, portanto $\frac{\partial \min}{\partial x}(x,x)$ não está definida. Verifique que $\frac{\partial \min}{\partial y}(x,x)$ não está definida. Se $h > 0$, então $\frac{\max(x+h,x)-\max(x,x)}{h} = 1$; se $h < 0$, então $\frac{\max(x+h,x)-\max(x,x)}{h} = 0$, portanto $\frac{\partial \max}{\partial x}(x,x)$ não está definida. Verifique que $\frac{\partial \max}{\partial y}(x,x)$ não está definida.

Mostremos que $\min(x,y)$ é côncava.

$$\begin{aligned} & \min((1-\lambda) \cdot (x_1, y_1) + \lambda \cdot (x_2, y_2)) = \\ & \min((1-\lambda) \cdot x_1 + \lambda \cdot x_2, (1-\lambda) \cdot y_1 + \lambda \cdot y_2) = \\ & \frac{(1-\lambda) \cdot x_1 + \lambda \cdot x_2 + (1-\lambda) \cdot y_1 + \lambda \cdot y_2 - \left| ((1-\lambda) \cdot x_1 + \lambda \cdot x_2) - ((1-\lambda) \cdot y_1 + \lambda \cdot y_2) \right|}{2} = \\ & \frac{(1-\lambda) \cdot (x_1 + y_1) + \lambda \cdot (x_2 + y_2) - \left| (1-\lambda) \cdot (x_1 - y_1) + \lambda \cdot (x_2 - y_2) \right|}{2} \geq \\ & \frac{(1-\lambda) \cdot ((x_1 + y_1) - |x_1 - y_1|) + \lambda \cdot ((x_2 + y_2) - |x_2 - y_2|)}{2} \geq \\ & (1-\lambda) \cdot \min(x_1, y_1) + \lambda \cdot \min(x_2, y_2), 0 < \lambda < 1. \end{aligned}$$

Observe que usamos a desigualdade $|a+b| \leq |a|+|b|$ na passagem da antepenúltima para a penúltima etapa. Analogamente, mostramos que $\max(x,y)$ é convexa ou, simplesmente, observamos que f é convexa se e só se $-f$ é côncava.

Exemplo 5: As definições dadas acima não são equivalentes. Em relação à concavidade, por exemplo, não é verdade que $f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2)$, $\forall x_1, x_2 \in S, \lambda \in [0,1]$ implique $f(x) \leq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, $\forall x$, pelo fato de que a função pode não ter planos tangentes em alguns pontos; é possível que as derivadas parciais de f não estejam definidas ou, se definidas, não sejam contínuas.

4. Pontos críticos de funções convexas e côncavas

Digamos que $f: S \rightarrow \mathbb{R}^n$ seja **convexa**, então $f(x) \geq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, $\forall x^*, x \in S$. Se x^* é um ponto crítico de f , então $\nabla f(x^*) = 0$, portanto, $f(x) \geq f(x^*)$, $x \in S$. Logo, x^* é um ponto de mínimo global de f em S . Da mesma maneira, digamos que $f: S \rightarrow \mathbb{R}^n$ seja **côncava**, então $f(x) \leq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, $\forall x^*, x \in S$. Se x^* é um ponto crítico de f , então $f(x) \leq f(x^*)$, $x \in S$ e x^* é um ponto de máximo global de f em S .

Se $f: S \rightarrow \mathbb{R}^n$ é estritamente convexa ou estritamente côncava e se f tiver ponto crítico, então esse ponto crítico é único. Nesses casos, f tem um único ponto de mínimo global se for convexa e um único ponto de máximo global se for côncava.

5. Funções quase-côncavas e funções quase-convexas.

Digamos que x^* seja um ponto crítico do lagrangeano da função $f(x)$ sujeita às restrições $g(x) \leq a$ e $h(x) \geq b$ e que as duas fronteiras sejam ativas em x^* , isto é, $g(x^*) = a$ e $h(x^*) = b$. Sabemos que $\nabla f(x^*) = \lambda^* \cdot \nabla g(x^*) + \mu^* \cdot \nabla h(x^*)$, o que implica que $\nabla f(x^*) \cdot (x - x^*) = \lambda^* \cdot \nabla g(x^*) \cdot (x - x^*) + \mu^* \cdot \nabla h(x^*) \cdot (x - x^*)$. Se x satisfaz as restrições, então $g(x) \leq a = g(x^*)$ e $h(x) \geq b = h(x^*)$, ou seja, $g(x) - g(x^*) \leq 0$ e $h(x) - h(x^*) \geq 0$. Digamos que $g(x)$ seja uma função convexa e que $h(x)$ seja uma função côncava. Portanto, $0 \geq g(x) - g(x^*) \geq \nabla g(x^*) \cdot (x - x^*)$ e $0 \leq h(x) - h(x^*) \leq \nabla h(x^*) \cdot (x - x^*)$. Digamos que x^* satisfaça as condições para um máximo, então $\lambda^* \geq 0$ e $\mu^* \leq 0$. Portanto, $\lambda^* \cdot \nabla g(x^*) \cdot (x - x^*) \leq 0$ e $\mu^* \cdot \nabla h(x^*) \cdot (x - x^*) \leq 0$. Logo, $\nabla f(x^*) \cdot (x - x^*) \leq 0$. Se supomos que $f(x)$ seja côncava, então $f(x) - f(x^*) \leq \nabla f(x^*) \cdot (x - x^*) \leq 0$ e, portanto, x^* é um máximo global para $f(x)$, dadas as restrições $g(x) \leq a$ e $h(x) \geq b$. Nesse caso, as condições de Kuhn-Tucker tornaram-se suficientes para assegurar que x^* é um máximo.

Igualmente, podemos mostrar que se $f(x)$ é convexa e x^* satisfaz as condições de Kuhn-Tucker para um mínimo, então x^* é mínimo global, pois, nesse caso, $\lambda^* \cdot \nabla g(x^*) \cdot (x - x^*) \geq 0$ porque $\lambda^* \leq 0$ e $\mu^* \cdot \nabla h(x^*) \cdot (x - x^*) \geq 0$ porque $\mu^* \geq 0$. Logo, $f(x) - f(x^*) \geq \nabla f(x^*) \cdot (x - x^*) \geq 0$.

Observe que podemos generalizar facilmente os resultados para qualquer número de fronteiras, sendo algumas ativas e outras inativas, pois quando as fronteiras são inativas os respectivos multiplicadores são nulos e, portanto, podemos desconsiderá-los na argumentação acima.

Portanto, temos o seguinte resultado: seja o problema:

$$\begin{array}{ll} \max & f(x) \\ \text{s.t.} & g_1(x) \leq a_1; \dots; g_k(x) \leq a_k; \\ & h_1(x) \geq b_1; \dots; h_l(x) \geq b_l \end{array}$$

Suponhamos que a função $f(x)$ seja côncava, que as $g_1(x), \dots, g_k(x)$ sejam convexas e que as $h_1(x), \dots, h_l(x)$ sejam côncavas, então se x^* satisfaz as condições de Kuhn-Tucker para um máximo, então x^* é máximo global na região determinada pelas restrições.

De maneira análoga, se a função $f(x)$ é convexa e x^* satisfaz as condições de Kuhn-Tucker para um mínimo, então x^* é mínimo global na região determinada pelas restrições.

Observe que, caso as restrições sejam funções afins, as exigências de convexidade e concavidade das restrições estão satisfeitas. Nesses casos, a exigência de concavidade ou convexidade incide apenas sobre a função f . Nos problemas de programação linear, a função f também é afim. Nesses casos, basta observar que os multiplicadores de Lagrange são maiores ou iguais a zero, para assegurar que estamos no mínimo ou no máximo global do problema e que os respectivos multiplicadores são as correspondentes soluções do problema primal ou do problema dual.

Se analisamos a argumentação anterior, observamos que o fundamental está nas inferências: i) se $g(x) - g(x^*) \leq 0$, então $\nabla g(x^*) \cdot (x - x^*) \leq 0$, pois, pela convexidade da função $g(x)$, temos $0 \geq g(x) - g(x^*) \geq \nabla g(x^*) \cdot (x - x^*)$; ii) se $0 \leq h(x) - h(x^*)$, então $0 \leq \nabla h(x^*) \cdot (x - x^*)$, pois pela concavidade da função $h(x)$, temos $0 \leq h(x) - h(x^*) \leq \nabla h(x^*) \cdot (x - x^*)$. É fundamental perceber que as condições:

i) se $g(x) - g(x^*) \leq 0$, então $\nabla g(x^*) \cdot (x - x^*) \leq 0$;

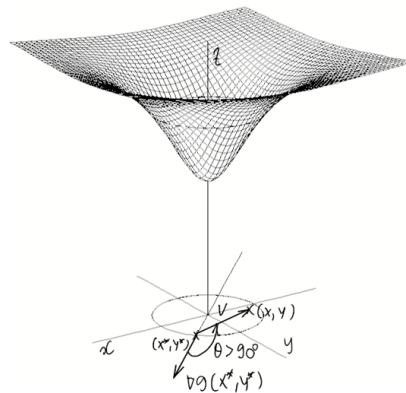
ii) se $0 \leq h(x) - h(x^*)$, então $0 \leq \nabla h(x^*) \cdot (x - x^*)$,

não se referem à relação entre os gráficos das funções e os respectivos planos tangentes, mas sim são consequências dessas relações. Portanto, se exigirmos que as funções $f(x)$ e as funções $h_1(x), \dots, h_l(x)$ satisfaçam a condição ii), e que as funções $g_1(x), \dots, g_k(x)$ satisfaçam

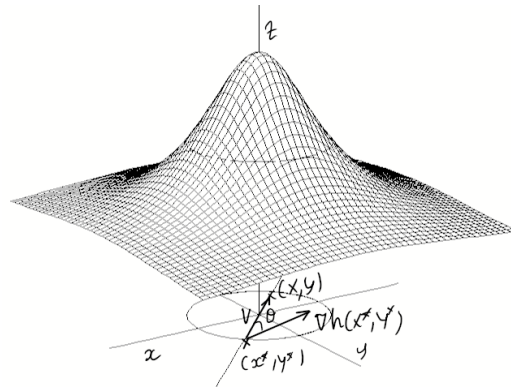
a condição i), obtemos os mesmos resultados em relação aos mínimos e máximos. As condições i) e ii) são um pouco mais gerais, na medida em que as funções convexas e côncavas as satisfazem. De fato, são condições satisfeitas por funções que não são nem convexas e nem côncavas.

Exemplo 6: Seja $g(x,y) = 2 - \frac{1}{1+x^2+y^2}$. Observe que g está satisfazendo a condição se $g(x,y) - g(x^*,y^*) \leq 0$, então $\nabla g(x^*,y^*) \cdot (x-x^*, y-y^*) \leq 0$. A circunferência é a isoquanta, digamos $g=a$, o ponto (x^*,y^*) é tal que $g(x^*,y^*)=a$, o ponto (x,y) é tal que está no interior do círculo $g(x,y) \leq a = g(x^*,y^*)$ e o ângulo θ , que o vetor $v=(x-x^*, y-y^*)$ faz com o gradiente $\nabla g(x^*,y^*)$, é maior que 90° . Não é muito difícil, observando o gráfico, concluir que essa função não é nem convexa, nem côncava.

Se $h(x,y) = \frac{1}{2} + \frac{1}{1+x^2+y^2}$, então se h satisfaz a condição $h(x,y) - h(x^*,y^*) \geq 0$, então $\nabla h(x^*,y^*) \cdot (x-x^*, y-y^*) \geq 0$. A circunferência é a isoquanta, digamos $h=b$, o ponto (x^*,y^*) é tal que $h(x^*,y^*)=b$, o ponto (x,y) é tal que está no interior do círculo $h(x,y) \geq b = h(x^*,y^*)$ e o ângulo θ , que o vetor $v=(x-x^*, y-y^*)$ faz com o gradiente $\nabla h(x^*,y^*)$, é menor que 90° . Novamente, observando o gráfico, concluímos que essa função não é nem convexa, nem côncava. Observe os desenhos abaixo:



A $g(x,y) = 2 - \frac{1}{1+x^2+y^2}$ é exemplo de uma função que dizemos “quase-convexa”:



A função $h(x, y) = \frac{1}{2 + \sqrt{1 + x^2 + y^2}}$ é exemplo de uma função que dizemos “quase-côncava”.

Portanto, no Teorema de Kuhn-Tucker acima, se substituirmos as condições sobre as funções serem convexas ou côncavas por quase-convexas ou quase-côncavas, obtemos o mesmo resultado.

As definições de quase-convexa e quase-côncava são: diz-se que $f: S \rightarrow \mathbb{R}^n$ é **quase-convexa** se

$$f(x) \leq f(x^*) \rightarrow \nabla f(x^*) \cdot (x - x^*) \leq 0;$$

diz-se que $f: S \rightarrow \mathbb{R}^n$ é **quase-côncava** se

$$f(x) \geq f(x^*) \rightarrow \nabla f(x^*) \cdot (x - x^*) \geq 0.$$

Para intuirmos a ideia geométrica da quase-convexidade e da quase-concavidade, definimos os chamados **contornos inferiores e superiores** de uma função. O **contorno inferior de nível c** é definido como:

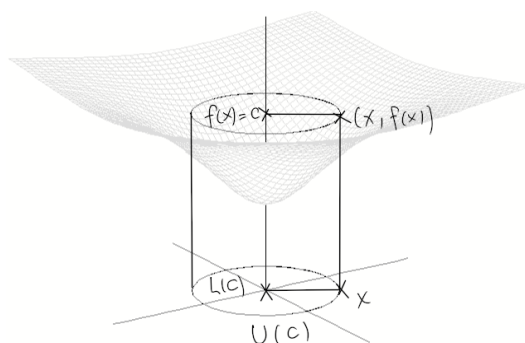
$$L(c) = \{x \in S \mid f(x) \leq c\},$$

ou seja, o contorno inferior de nível c é formado pela união de todas as isoquantas de nível menor ou igual a c . O **contorno superior de nível c** é definido como:

$$U(c) = \{x \in S \mid f(x) \geq c\},$$

ou seja, o contorno superior de nível c é formado pela união de todas as isoquantas de nível maior ou igual a c . Abaixo estão desenhadas duas funções.

O desenho abaixo representa uma função quase-convexa. Observe que o contorno superior está no exterior da circunferência e o contorno inferior é constituído pelos pontos interiores à circunferência. A circunferência é a isoquanta $f(x)=c$ de nível c . Podemos observar que, neste caso, o contorno superior $L(c) = \{x \in S \mid f(x) \leq c\}$ é um conjunto convexo e esta é uma definição alternativa para a quase-concavidade: $f: S \rightarrow \mathbb{R}^n$ é quase-convexa se os contornos superiores $L(c)$ são conjuntos convexos.:

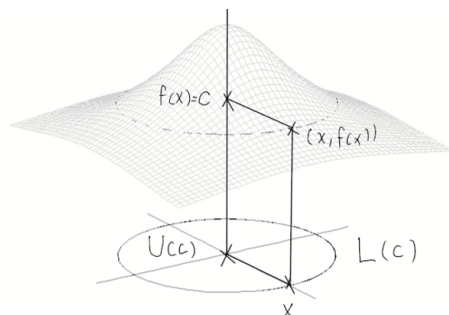


Tomemos dois pontos $x_1, x_2 \in S$, então $f(x_1) \leq f(x_2)$ ou $f(x_1) \geq f(x_2)$. Digamos que $f(x_1) \leq f(x_2)$, então $x_1, x_2 \in L(f(x_2)) = \{x \in S \mid f(x) \leq f(x_2)\}$, que é um conjunto convexo, então $f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq f(x_2) = \max(f(x_1), f(x_2))$, $\lambda \in [0,1]$. Portanto, temos uma nova definição: $f: S \rightarrow \mathbb{R}^n$ diz-se que é **quase-convexa** se

$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq \max(f(x_1), f(x_2)), \forall x_1, x_2 \in S, \lambda \in [0,1].$$

O desenho abaixo representa uma função quase-côncava. Observe que o contorno superior está no interior da circunferência e o contorno inferior é constituído pelos pontos externos à circunferência. A circunferência é a isoquanta $f(x)=c$ de nível c .

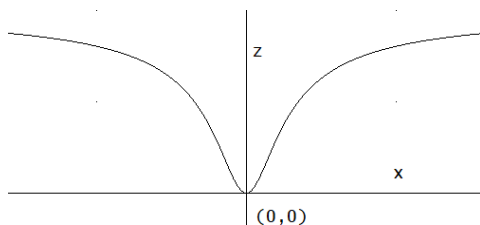
Podemos observar que, neste caso, o contorno superior $U(c) = \{x \in S \mid f(x) \geq c\}$ é um conjunto convexo e esta é uma definição alternativa para a quase-concavidade: $f: S \rightarrow \mathbb{R}^n$ é quase-côncava se os contornos superiores $U(c)$ são conjuntos convexos:



Tomemos dois pontos $x_1, x_2 \in S$, então $f(x_1) \leq f(x_2)$ ou $f(x_1) \geq f(x_2)$. Digamos que $f(x_1) \geq f(x_2)$, então $x_1, x_2 \in U(f(x_2)) = \{x \in S \mid f(x) \geq f(x_2)\}$, que é um conjunto convexo, então $f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \geq f(x_2) = \min(f(x_1), f(x_2)), \lambda \in [0,1]$. Portanto, temos uma nova definição: $f: S \rightarrow \mathbb{R}^n$ diz-se que é **quase-côncava** se

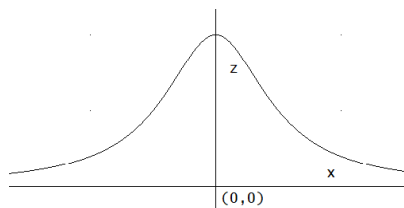
$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \geq \min(f(x_1), f(x_2)), \forall x_1, x_2 \in S, \lambda \in [0,1].$$

Exemplo 7: Seja $f(x,y) = 1 - \frac{1}{1+x^2+y^2}$ não é convexa e nem côncava. Por exemplo, se fazemos $y=0$, obtemos $f(x,0) = 1 - \frac{1}{1+x^2}$, cujo gráfico está abaixo e é o perfil da função f sobre o eixo x . Este perfil não representa nem uma função convexa nem uma função côncava.



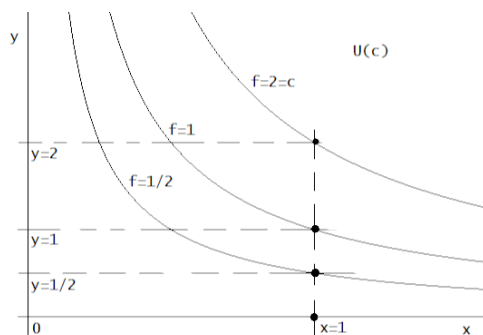
Mas, f é quase-convexa, pois as isoquantas $f(x,y) = c$ são circunferências $x^2 + y^2 = \frac{c}{1-c}$ de raio $r = \sqrt{\frac{c}{1-c}}$, para $0 \leq c < 1$. Se c aumenta, então $r = \sqrt{\frac{c}{1-c}}$ aumenta, portanto, o contorno inferior de nível c é o círculo de raio $r = \sqrt{\frac{c}{1-c}}$.

Exemplo 8: Seja $f(x,y) = \frac{1}{1+x^2+y^2}$ não é convexa e nem côncava. Por exemplo, se fazemos $y=0$, obtemos $f(x,0) = \frac{1}{1+x^2}$, cujo gráfico está abaixo e é o perfil da função f sobre o eixo x . Este perfil não representa nem uma função convexa nem uma função côncava.



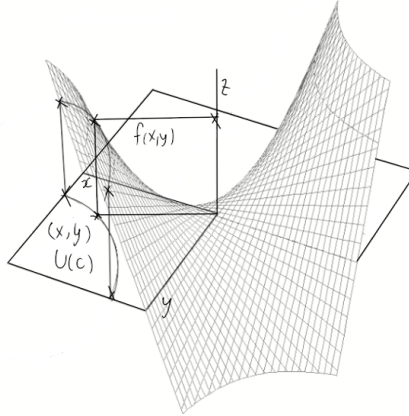
Mas, f é quase-côncava, pois as isoquantas $f(x,y)=c$ são circunferências $x^2+y^2 = \frac{1}{c}-1$ de raio $r = \sqrt{\frac{1}{c}-1}$, para $0 < c \leq 1$. Se c aumenta, então $r = \sqrt{\frac{1}{c}-1}$ diminui, portanto, o contorno superior de nível c é o círculo de raio $r = \sqrt{\frac{1}{c}-1}$.

Exemplo 9: Consideremos a clássica sela $f(x,y) = x \cdot y$, apenas para $x > 0$ e $y > 0$. Claramente, $f(x,y) = x \cdot y$ para $x > 0$ e $y > 0$ é quase-côncava. A isoquanta de nível $c > 0$ é $y = \frac{c}{x}$.

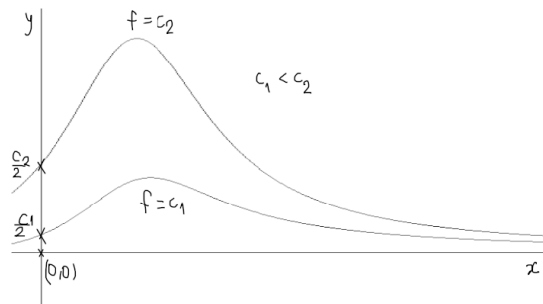


Se $x=y$, temos a parábola convexa $f(x,x) = x^2$ e se fazemos, por exemplo, $y=5-x$, resulta na parábola côncava $f(x,5-x) = x \cdot (5-x)$. Essas duas curvas claramente comprovam que $f(x,y) = x \cdot y$ não é nem côncava, nem convexa.

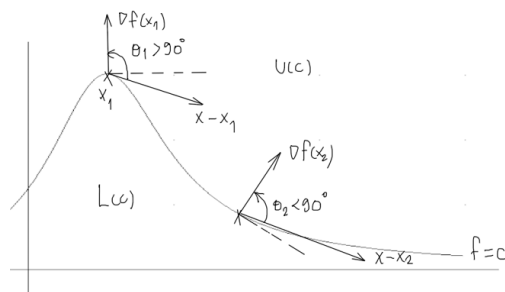
$f(x,y) = x \cdot y$ é sela girada 45° no sentido anti-horário. Como se pode observar no desenho acima.



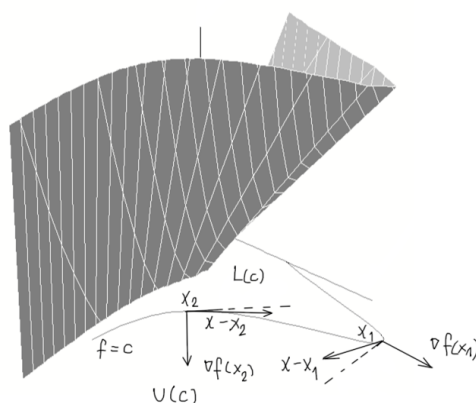
Exemplo 10: Uma função pode não ser nem quase-convexa, nem quase-côncava. Seja a função $f(x,y) = x^2 \cdot y - 2 \cdot x \cdot y - x + 2 \cdot y$ definida em $x > 0, y > 0$. Se numa isoquanta definirmos y em função de x , obtemos $y = \frac{x+c}{x^2-2 \cdot x+2}$, que estão desenhadas abaixo:



Essas curvas não são nem convexas, nem côncavas. Observe no desenho abaixo que nem os contornos inferiores, nem os contornos superiores são conjuntos convexas do $\mathbb{R}^2$. Como consequência, existem $x_1, x \in \mathbb{R}^2$, tal que $f(x) \geq f(x_1) = c$, ou seja, $x \in U(c)$, e $\nabla f(x_1) \cdot (x - x_1) < 0$, isto é, o ângulo entre $\nabla f(x_1)$ e $(x - x_1)$ é maior que 90° e existem $x_2, x \in \mathbb{R}^2$, tal que $f(x) \leq f(x_2) = c$, ou seja, $x \in L(c)$, e $\nabla f(x_2) \cdot (x - x_2) > 0$, isto é, o ângulo entre $\nabla f(x_2)$ e $(x - x_2)$ é menor que 90° .



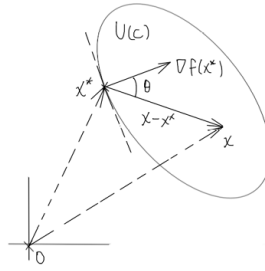
Abaixo, uma representação “tridimensional” do gráfico da função, da isoquanta de nível c e dos respectivos contornos:



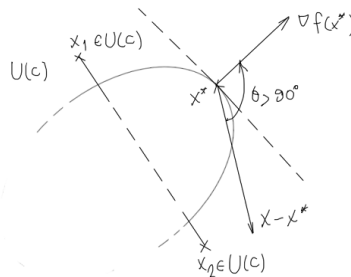
Teorema 1

- $f: S \rightarrow \mathbb{R}^n$ é quase-côncava se, e só se, $f(x) \geq f(x^*) \rightarrow \nabla f(x^*) \cdot (x-x^*) \geq 0$;
- $f: S \rightarrow \mathbb{R}^n$ é quase-convexa se, e só se, $f(x) \leq f(x^*) \rightarrow \nabla f(x^*) \cdot (x-x^*) \leq 0$.

Prova: Suponhamos que $f: S \rightarrow \mathbb{R}^n$ é quase-côncava, então qualquer contorno superior $U(c)$ de f é um conjunto convexo. Seja $f(x) \geq f(x^*)$, isto é, $x \in U(c)$, então, como vemos no desenho, necessariamente o ângulo entre $\nabla f(x^*)$ e $(x-x^*)$ é no máximo 90° , caso a isoquanta $f=c$ se confunda com a perpendicular ao gradiente $\nabla f(x^*)$, então $\nabla f(x^*) \cdot (x-x^*) \geq 0$:

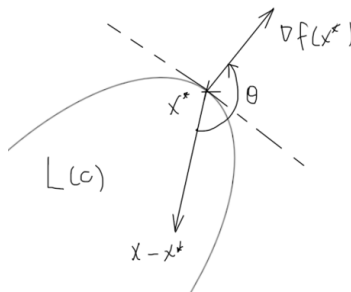


Por outro lado, se $f(x) \geq f(x^*)$ implica que $\nabla f(x^*) \cdot (x - x^*) \geq 0$, então facilmente nos convencemos que qualquer contorno superior $U(c)$ de f não pode ter a forma abaixo, em que $x^* \in U(c)$ $x_1 \in U(c)$ e $x_2 \in U(c)$:

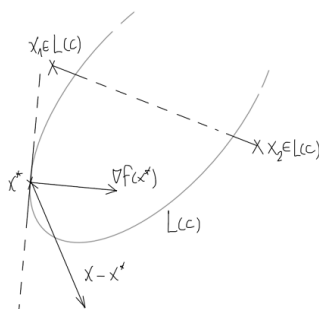


Observe que o segmento de reta $[x_1, x_2]$, unindo $x_1 \in U(c)$ e $x_2 \in U(c)$, não está inteiramente em $U(c)$. Se existem pontos $x \in U(c)$, tais que $\nabla f(x^*) \cdot (x - x^*) < 0$, então sempre existem pontos $x_1 \in U(c)$ e $x_2 \in U(c)$, tais que o segmento $[x_1, x_2]$ não está contido em $U(c)$. Logo, se $f(x) \geq f(x^*) \rightarrow \nabla f(x^*) \cdot (x - x^*) \geq 0$, então $U(c)$ é um conjunto convexo.

Para o caso de uma função quase-convexa, observe os desenhos abaixo. Suponhamos que qualquer contorno inferior $L(c)$ de f é um conjunto convexo. Seja $f(x) \leq f(x^*)$, isto é, $x \in L(c)$, então, como vemos no desenho, necessariamente o ângulo entre $\nabla f(x^*)$ e $(x - x^*)$ é no mínimo 90° , caso a isoquanta $f=c$ se confunda com a perpendicular ao gradiente $\nabla f(x^*)$, então $\nabla f(x^*) \cdot (x - x^*) \leq 0$:

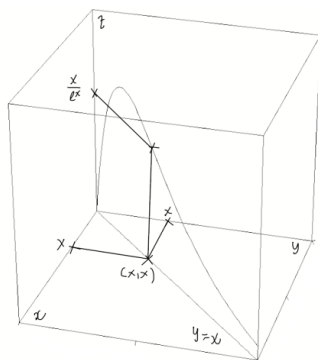


Por outro lado, se $f(x) \leq f(x^*)$ implica que $\nabla f(x^*) \cdot (x - x^*) \leq 0$, então facilmente nos convencemos que qualquer contorno inferior $L(c)$ de f não pode ter a forma abaixo, em que $x^* \in L(c)$, $x_1 \in L(c)$ e $x_2 \in L(c)$. Se existem pontos $x \in L(c)$, tais que $\nabla f(x^*) \cdot (x - x^*) > 0$, então sempre existem pontos $x_1 \in L(c)$ e $x_2 \in L(c)$, tais que o segmento $[x_1, x_2]$ não está contido em $L(c)$. Logo, se $f(x) \leq f(x^*) \rightarrow \nabla f(x^*) \cdot (x - x^*) \leq 0$, então $L(c)$ é um conjunto convexo.

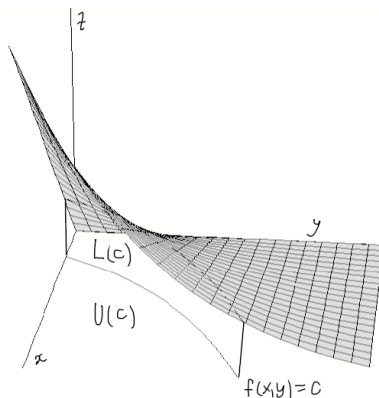


Exercício 2

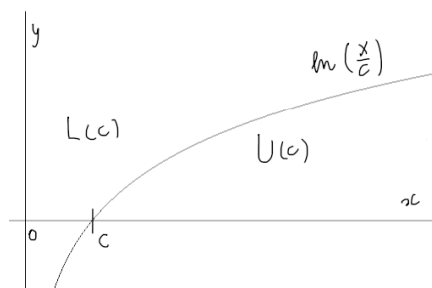
Considere a função $f(x, y) = \frac{x}{e^y}$ definida para $x \geq 0$ e $y \geq 0$. No gráfico abaixo está plotada a curva $(x, x, f(x, x)) = (x, x, \frac{x}{e^x})$. Observando este gráfico, conclua que f não é côncava e não é convexa:



No gráfico abaixo estão plotados o gráfico de $f(x, y) = \frac{x}{e^y}$ e uma isoquanta $f(x, y) = c$. Além disso, indicamos os contornos inferiores e superiores:



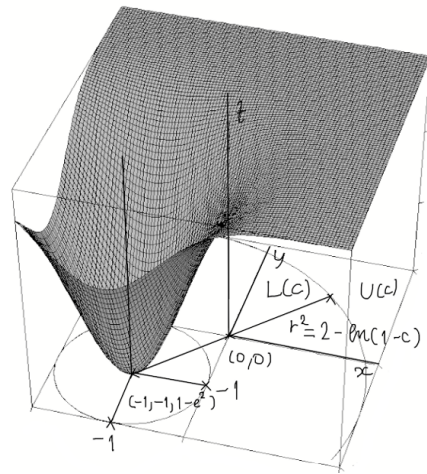
Mostre que a isoquanta $f(x, y) = c$ é tal que $y = \ln\left(\frac{x}{c}\right)$ e conclua que $f(x, y)$ é quase-côncava. Observe o gráfico abaixo:



Exercício 3

Mário Henrique Simonsen, no volume I do livro “Microeconomia”¹⁹, define a “utilidade” $u(x_1, x_2) = 1 - e^{2 - (x_1 + 1)^2 - (x_2 + 1)^2}$ para $x_1 > 0$ e $x_2 > 0$, que analisaremos mais à frente. Mostre que as isoquantas $u(x_1, x_2) = c$ são as circunferências $(x_1 + 1)^2 + (x_2 + 1)^2 = 2 - \ln(1 - c)$ centradas em $(-1, -1)$, de raio $r = \sqrt{2 - \ln(1 - c)}$, para $1 - e^2 < c < 1$. Em $(-1, -1)$, a função u tem o mínimo global $u(-1, -1) = 1 - e^2$. Observe que $r(c) = \sqrt{2 - \ln(1 - c)}$ é uma função crescente em c , pois $r'(c) = \frac{1}{2 \cdot (1 - c) \cdot \sqrt{2 - \ln(1 - c)}} > 0$. Portanto, se c cresce, os raios das circunferências crescem, implicando que os contornos inferiores estão nos interiores

dos círculos. Conclua desta observação que $u(x_1, x_2) = 1 - e^{2-(x_1+1)^2-(x_2+1)^2}$ é uma função quase-convexa.



6. Teorema de Kuhn-Tucker restrito a funções quase-côncavas e quase-convexas

Seja o problema:

$$\begin{array}{ll} \min/\max & f(x_1, \dots, x_n) \\ \text{sa} & g_1(x_1, \dots, x_n) \leq a_1; \\ & \vdots \\ & g_k(x_1, \dots, x_n) \leq a_k; \\ & h_1(x_1, \dots, x_n) \geq b_1; \\ & \vdots \\ & h_l(x_1, \dots, x_n) \geq b_l. \end{array}$$

Suponhamos que as funções $g_1(x), \dots, g_k(x)$ sejam **quase-convexas** e que as funções $h_1(x), \dots, h_l(x)$ sejam **quase-côncavas**.

i) Então, se a função $f(x)$ é **quase-convexa** e x^* satisfaz as condições de Kuhn-Tucker para um mínimo, então x^* é mínimo global na região determinada pelas restrições;

ii) Se a função $f(x)$ é **quase-côncava** e x^* satisfaz as condições de Kuhn-Tucker para um máximo, então x^* é máximo global na região determinada pelas restrições.

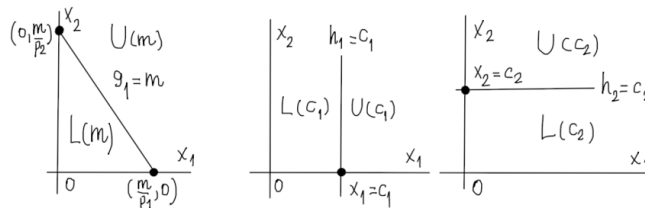
Esse resultado é facilmente estabelecido a partir dos comentários da seção 5. Demonstramos o caso ii). Digamos que x^* seja um ponto crítico do lagrangeano da função $f(x)$, sujeita às restrições $g_1(x) \leq a_1, \dots, g_k(x) \leq a_k$ e $h_1(x) \geq b_1, \dots, h_l(x) \geq b_l$, e suponhamos que apenas duas fronteiras sejam ativas em x^* , isto é, $g_i(x^*) = a_i$ e $h_j(x^*) = b_j$. Como $\nabla f(x^*) = \lambda_i^* \cdot \nabla g_i(x^*) + \mu_j^* \cdot \nabla h_j(x^*)$, então $\nabla f(x^*) \cdot (x - x^*) = \lambda_i^* \cdot \nabla g_i(x^*) \cdot (x - x^*) + \mu_j^* \cdot \nabla h_j(x^*) \cdot (x - x^*)$. Se x satisfaz as restrições, então $g_i(x) \leq a_i = g_i(x^*)$ e $h_j(x) \geq b_j = h_j(x^*)$, ou seja, $g_i(x) - g_i(x^*) \leq 0$ e $h_j(x) - h_j(x^*) \geq 0$. Digamos que $g_i(x)$ seja uma função quase-convexa e que $h_j(x)$ seja uma função quase-côncava. Portanto, $0 \geq \nabla g_i(x^*) \cdot (x - x^*)$ e $0 \leq \nabla h_j(x^*) \cdot (x - x^*)$. Digamos que x^* satisfaça as condições para um máximo, então $\lambda^* \geq 0$ e $\mu^* \leq 0$. Logo, $\lambda^* \cdot \nabla g_i(x^*) \cdot (x - x^*) \leq 0$ e $\mu^* \cdot \nabla h_j(x^*) \cdot (x - x^*) \leq 0$. Portanto, $\nabla f(x^*) \cdot (x - x^*) \leq 0$. Se supomos que $f(x)$ seja quase-côncava, então $f(x) - f(x^*) \leq 0$ e x^* é um máximo global para $f(x)$, dadas as restrições. O caso i) deixaremos como exercício.

Exemplo 11: Seja o problema:

$$\begin{aligned} \max \quad & f(x_1, x_2) = x_1 \cdot x_2 \\ \text{sa} \quad & g(x_1, x_2) = p_1 \cdot x_1 + p_2 \cdot x_2 \leq m; \\ & h_1(x_1, x_2) = x_1 \geq 0; \\ & h_2(x_1, x_2) = x_2 \geq 0. \end{aligned}$$

A função objetivo $f(x_1, x_2) = x_1 \cdot x_2$ é quase-côncava. Veja o gráfico abaixo.

As funções $g(x_1, x_2) = p_1 \cdot x_1 + p_2 \cdot x_2$, $h_1(x_1, x_2) = x_1$ e $h_2(x_1, x_2) = x_2$ são lineares e, portanto, trivialmente quase-côncavas e quase-convexas simultaneamente. Veja os gráficos abaixo.:



Como as isoquantas de $f(x_1, x_2) = x_1 \cdot x_2$ não cortam os eixos coordenados, então as fronteiras $h_1(x_1, x_2) = x_1$ e $h_2(x_1, x_2) = x_2$ são sempre inativas e as condições de Kuhn-Tucker para um máximo se reduzem a $\frac{\partial f}{\partial x_1}(x_1, x_2) - \lambda \cdot \frac{\partial g}{\partial x_1}(x_1, x_2) = x_2 - \lambda \cdot p_1 = 0$,

$\frac{\partial f}{\partial x_2}(x_1, x_2) - \lambda \cdot \frac{\partial g}{\partial x_2}(x_1, x_2) = x_1 - \lambda \cdot p_2 = 0$ e $\lambda > 0$. Pelo Teorema de Kuhn-Tucker sujeito a quase-concavidade e quase-convexidade, essas condições são suficientes para assegurar que o ponto crítico $(x_1^*, x_2^*) = \left(\frac{m}{2 \cdot p_1}, \frac{m}{2 \cdot p_2}\right)$, $\lambda^* = \frac{m}{2 \cdot p_1 \cdot p_2}$, é um ponto de máximo global de $f(x_1, x_2) = x_1 \cdot x_2$, dadas as restrições acima.

Quando as restrições são lineares, o que é uma condição comum, se a função objetivo for quase-côncava, se um ponto, ou pontos, satisfaz as condições de Kuhn-Tucker para máximo, então, é máximo global; se satisfaz as condições de Kuhn-Tucker para um mínimo, é mínimo global. Esse é o caso da maximização de uma utilidade restrita às restrições orçamentárias. Portanto, a exigência posta sobre as funções de utilidades: exige-se que sejam quase-côncavas, como no exemplo acima $f(x_1, x_2) = x_1 \cdot x_2$. Uma última observação sobre $f(x_1, x_2) = x_1 \cdot x_2$. Essa função é a sela usual girada 45° no sentido anti-horário, como já observamos no exemplo.

Exemplo 12: Seja o problema:

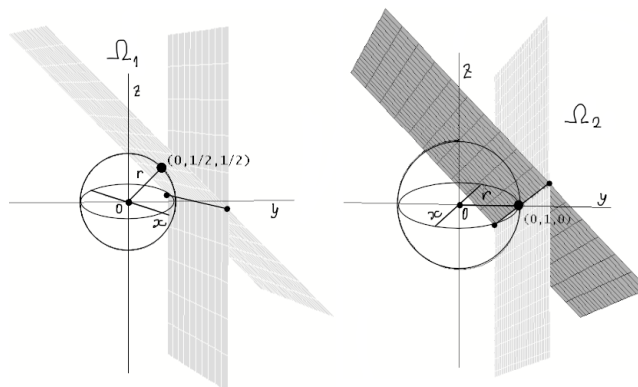
$$\begin{aligned} \min \quad & f(x_1, x_2) = w_1 \cdot x_1 + w_2 \cdot x_2 \\ \text{sa} \quad & h(x_1, x_2) = x_1 \cdot x_2 \geq y; \\ & h_1(x_1, x_2) = x_1 \geq 0; \\ & h_2(x_1, x_2) = x_2 \geq 0. \end{aligned}$$

Observe que $f(x_1, x_2) = w_1 \cdot x_1 + w_2 \cdot x_2$ é linear, portanto quase-convexa, e que as restrições $h(x_1, x_2) = x_1 \cdot x_2$, $h_1(x_1, x_2) = x_1$ e $h_2(x_1, x_2) = x_2$ são quase-côncavas. Logo, os pontos que satisfizerem as condições de Kuhn-Tucker para um mínimo são mínimos globais. No caso, $(x_1^*, x_2^*) = \left(\sqrt{\frac{w_2 \cdot y}{w_1}}, \sqrt{\frac{w_1 \cdot y}{w_2}}\right)$ e o multiplicador associado à restrição $h(x_1, x_2) = x_1 \cdot x_2 \geq y$ é $\mu^* = \sqrt{\frac{w_1 \cdot w_2}{y}} > 0$. Desse modo, $\left(\sqrt{\frac{w_2 \cdot y}{w_1}}, \sqrt{\frac{w_1 \cdot y}{w_2}}\right)$ é um mínimo global. Observe que para que um ponto (x_1^*, x_2^*) , $\mu^* > 0$, seja um mínimo global, basta exigir $h(x_1, x_2)$ quase-côncava. Essa é a razão pela qual exigimos que também as funções de produção, tanto quanto as de utilidade, sejam quase-côncavas.

Exemplo 13: Considere o problema de localizar o ponto que está mais próximo da origem:

$$\begin{array}{ll} \min & x^2 + y^2 + z^2 \\ \text{sa} & 0 \cdot x + y + 0 \cdot z \leq 1 \\ & 0 \cdot x + y + z \geq 1. \end{array}$$

$f(x, y, z) = x^2 + y^2 + z^2$ é convexa e, portanto, quase-convexa; $g(x, y, z) = y$ e $h(x, y, z) = y + z$ são lineares. Logo, o ponto que satisfizer as condições de Kuhn-Tucker para um mínimo é mínimo global. Se o lagrangeano é $L = x^2 + y^2 + z^2 - \lambda \cdot (y - 1) - \mu \cdot (y + z - 1)$, então $(0, 1/2, 1/2)$, $\lambda = 0$, $\mu = 1$, é o ponto mais próximo da origem que satisfaz as restrições que definem a região viável Ω . A distância do ponto à origem é $r = \frac{\sqrt{2}}{2}$.



Considere o problema de localizar o ponto que está mais próximo da origem:

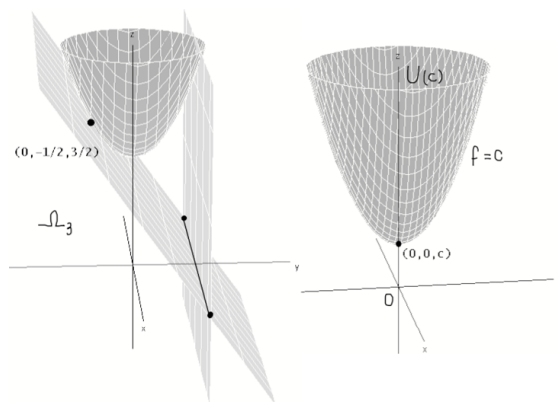
$$\begin{array}{ll} \min & x^2 + y^2 + z^2 \\ \text{sa} & 0 \cdot x + y + 0 \cdot z \geq 1 \\ & 0 \cdot x + y + z \geq 1. \end{array}$$

O lagrangeano é $\mathcal{L} = x^2 + y^2 + z^2 - \mu_1 \cdot (y - 1) - \mu_2 \cdot (y + z - 1)$, então $(0, 1, 0)$, $\mu_1 = 1$, $\mu_2 = 0$, é o ponto mais próximo da origem que satisfaz as restrições que definem a região viável Ω_2 . A distância do ponto $(0, 1, 0)$ à origem é $r = 1$. A reta na interseção dos planos é a reta $x = t$, $y = 1$, $z = 0$. O ponto $(0, 1, 0)$ é o ponto sobre a reta mais próximo da origem. O gráfico da esquerda representa o primeiro problema e o gráfico da direita, o segundo problema.

Exemplo 14: Considere o problema de localizar o ponto que está mais próximo da origem:

$$\begin{array}{ll} \max & z - x^2 - y^2 \\ \text{sa} & 0 \cdot x + y + 0 \cdot z \leq 1 \\ & 0 \cdot x + y + z \leq 1. \end{array}$$

$f(x, y, z) = z - x^2 - y^2$ é quase-côncava, pois os contornos superiores $U(c)$ de f são os $U(c) = \{(x, y, z) \mid z \geq c + x^2 + y^2\}$, que são conjuntos convexos. Veja o gráfico abaixo. Logo, o ponto que satisfizer as condições de Kuhn-Tucker para um máximo é máximo global. Se o lagrangeano é $\mathcal{L} = z - x^2 - y^2 - \lambda_1 \cdot (y - 1) - \lambda_2 \cdot (y + z - 1)$, então $(0, -1/2, 3/2)$, $\lambda_1 = 0$, $\lambda_2 = 1$, é o ponto de máximo global de $f(x, y, z) = z - x^2 - y^2$ na região Ω_3 . O valor máximo de $f = \frac{5}{4}$.



Na verdade, $f(x, y, z) = z - x^2 - y^2$ é uma função côncava, pois $f(x, y, z) \leq f(x_0, y_0, z_0) + \nabla f(x_0, y_0, z_0) \cdot (x - x_0, y - y_0, z - z_0)$. Observe que $z - x^2 - y^2 \leq z_0 - x_0^2 - y_0^2 + (-2 \cdot x_0) \cdot (x - x_0) + (-2 \cdot y_0) \cdot (y - y_0) + 1 \cdot (z - z_0)$, para quaisquer (x_0, y_0, z_0) e (x, y, z) , pois $0 \leq (x - x_0)^2 + (y - y_0)^2$. Todavia, $f(x, y, z) = z - x^2 - y^2$ não é estritamente côncava, pois nos pontos (x_0, y_0, z) ,

$$\begin{aligned} f(x_0, y_0, z) &= f(x_0, y_0, z_0) + \nabla f(x_0, y_0, z_0) \cdot (x - x_0, y - y_0, z - z_0), \\ z - x_0^2 - y_0^2 &= z_0 - x_0^2 - y_0^2 + 1 \cdot (z - z_0), \end{aligned}$$

ou seja, f e o plano tangente se confundem. A reta $x = x_0, y = y_0, z = z_0 + t$, $w = z_0 - x_0^2 - y_0^2 + t$, que é uma reta no $\mathbb{R}^4$, está tanto sobre o gráfico de f , quanto sobre o plano tangente.

Exemplo 15: Considere o problema de localizar o ponto mais próximo da origem sobre o plano $a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z = b_1 \neq 0$. A origem não está sobre o plano; $(0,0,0)$ está no semiespaço $a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z < b_1$, se $b_1 > 0$, e está em $a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z > b_1$, se $b_1 < 0$. Digamos que se $b_1 > 0$, então o problema de localizar o ponto sobre o plano mais próximo da origem pode ser formulado como:

$$\begin{aligned} \min \quad & x^2 + y^2 + z^2 \\ \text{sa} \quad & a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z \geq b_1. \end{aligned}$$

Como $f(x, y, z) = x^2 + y^2 + z^2$ é convexa e $h(x, y, z) = a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z$ é linear, qualquer ponto $P(x, y, z)$ viável que satisfaça as condições de Kuhn-Tucker para um mínimo é mínimo global. A primeira condição de Kuhn-Tucker assegura que $\begin{pmatrix} 2 \cdot x^* \\ 2 \cdot y^* \\ 2 \cdot z^* \end{pmatrix} = \mu_1 \cdot \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix}$. Em outras palavras, se $2 \cdot \begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix} = 2 \cdot v^*$ e $a_1 = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix}$, então $2 \cdot v^* = \mu_1 \cdot a_1$. Como $a_1' \cdot v^* \geq b_1 > 0$, então $0 < 2 \cdot b_1 = 2 \cdot a_1' \cdot v^* = \mu_1 \cdot a_1' \cdot a_1 = \mu_1 \cdot \|a_1\|^2$, portanto $\mu_1 > 0$. Logo, $a_{11} \cdot x^* + a_{12} \cdot y^* + a_{13} \cdot z^* = b_1$ e $v^* = \begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix}$ é um mínimo global.

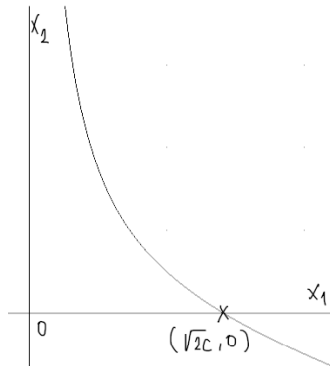
Se $b_1 < 0$, o problema pode ser formulado como:

$$\begin{aligned} \min \quad & x^2 + y^2 + z^2 \\ \text{sa} \quad & a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z \leq b_1. \end{aligned}$$

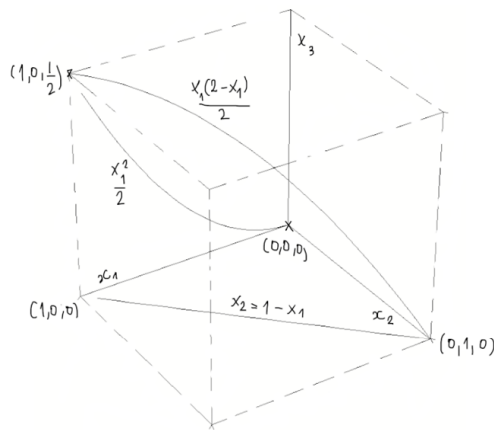
Nesse caso, $0 > 2 \cdot b_1 = 2 \cdot a_1' \cdot v^* = \lambda_1 \cdot a_1' \cdot a_1 = \lambda_1 \cdot \|a_1\|^2$, então $\lambda_1 < 0$. Portanto, v^* é um mínimo global.

Exercício 4

$u(x_1, x_2) = \frac{x_1^2}{2} + x_1 \cdot x_2$, $x_1 \geq 0, x_2 \geq 0$. Mostre que a isoquanta $u(x_1, x_2) = \frac{x_1^2}{2} + x_1 \cdot x_2 = c > 0$ pode ser representada por $x_2 = \frac{2 \cdot c - x_1^2}{2 \cdot x_1}$, para $x_1 > 0, x_2 \geq 0$. A isoquanta de nível zero, $u(x_1, x_2) = 0$, corresponde ao eixo $x_1 = 0$. Abaixo representamos as isoquantas $x_2 = \frac{2 \cdot c - x_1^2}{2 \cdot x_1}$, $c > 0$, no quadrante $x_1 \geq 0, x_2 \geq 0$. Observando este gráfico, salientado que $\frac{dx_2}{dx_1} < 0$, $\frac{d^2x_2}{dx_1^2} > 0$ e que a curva $x_2 = \frac{2 \cdot c - x_1^2}{2 \cdot x_1}$ corta o eixo $x_1 > 0$ no ponto $(\sqrt{2 \cdot c}, 0)$, conclua que u é uma função quase-côncava:



Observando as curvas $\left(x_1, 0, \frac{x_1^2}{2}\right)$ e as curvas $\left(x_1, 1-x_1, \frac{x_1 \cdot (2-x_1)^2}{2}\right)$, conclua que u não é nem côncava, nem convexa.



Exercício 5

Estudando os contornos inferiores e superiores, determine se as funções abaixo são quase-côncavas ou quase-convexas em

$$\mathbb{R}_+^2 = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0\};$$

i) $f(x, y) = x \cdot y + x$;

ii) $f(x, y) = \sqrt{2 \cdot x + 3 \cdot y}$;

iii) $f(x, y) = \ln(1 - (x+1)^2 - (y+1)^2)$;

iv) $f(x, y) = x \cdot y + x^2$;

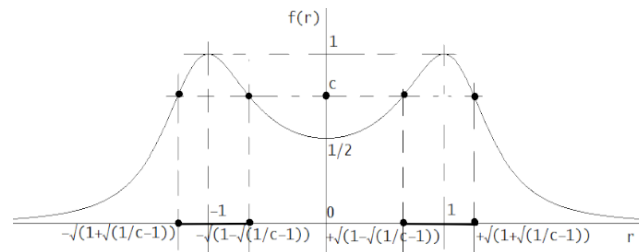
v) $f(x, y, z) = \ln((x+1)^2 + (y+2)^2 + z^2 + 1)$;

vi) $f(x,y) = \frac{1}{\sqrt{x^2+y^2}}$;

vii) $f(x,y) = \ln x + y + 1$.

Exercício 6

Considere a função $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(r) = \frac{1}{1+(1-r^2)^2}$, cujo gráfico está desenhado abaixo:



Se $\frac{1}{1+(1-r^2)^2} = c$, então $r = \pm \sqrt{1 \pm \sqrt{\frac{1}{c} - 1}}$. A partir disso, observe os contornos inferiores e superiores $f(r) = c < 1$: $L(c) = \left(-\infty, -\sqrt{1 + \sqrt{\frac{1}{c} - 1}}\right] \cup \left[\sqrt{1 + \sqrt{\frac{1}{c} - 1}}, +\infty\right)$ e $U(c) = \left[-\sqrt{1 + \sqrt{\frac{1}{c} - 1}}, -\sqrt{1 - \sqrt{\frac{1}{c} - 1}}\right] \cup \left[\sqrt{1 - \sqrt{\frac{1}{c} - 1}}, \sqrt{1 + \sqrt{\frac{1}{c} - 1}}\right]$. Observe o gráfico acima. Sem dúvida, $f(r) = \frac{1}{1+(1-r^2)^2}$ não é nem quase-côncava, nem quase-convexa.

Seja $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x,y) = \frac{1}{1+(1-x^2-y^2)^2}$. Determine os contornos $L(c)$ e $U(c)$ para f .

7. O resultado principal do artigo “Nonlinear Programming” (1950)

Na introdução do clássico artigo “Nonlinear Programming”, os autores esclarecem:

Este artigo formula condições necessárias e suficientes para um ponto de sela de qualquer função diferenciável $\phi(x,u)$ com argumentos não-negativos x e u (seção 2). Através de um lagrangeano $\phi(x,u)$, aplica-se este resultado a um máximo de uma função diferenciável $g(x)$ restrita por desigualdades envolvendo funções diferenciáveis $f_h(x)$, moderadamente qualificadas (seção 3). Então, mostra-se (seção 4) a equivalência acima entre um máximo restrito de $g(x)$ e um ponto de sela do Lagrangeano $\phi(x,u)$; requer-se apenas [para obter este resultado] que as funções $g(x)$ e $f_h(x)$ sejam côncavas (diferenciáveis) para x não-negativo.²⁰

Este resultado é o “verdadeiro” Teorema de Kuhn-Tucker que exporemos abaixo, seguindo a demonstração dos autores.

Observemos o lema simples:

Lema 1: Seja $f: [0, +\infty) \rightarrow \mathbb{R}$ derivável e $x_0 \in [0, +\infty)$:

i) se x_0 é ponto de máximo local de f , então (1) $f'(x_0) \leq 0$; (2) $f'(x_0) \cdot x_0 = 0$;

ii) se x_0 é ponto de mínimo local de f , então (1) $f'(x_0) \geq 0$; (2) $f'(x_0) \cdot x_0 = 0$.

Prova: o resultado é bastante evidente. Mostremos i). Se $x_0 = 0$ é máximo local, então, para $x > 0$, temos $\frac{f(x) - f(0)}{x} \leq 0$, logo $f'(0) = \lim_{x \rightarrow 0^+} \frac{f(x) - f(0)}{x} \leq 0$. Se $x_0 > 0$, então $f'(x_0) = 0$, o que implica $f'(x_0) \cdot x_0 = 0$. O caso ii) é análogo.

Seja uma função $\phi: \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow \mathbb{R}$. Vamos notar por $\nabla_x \phi(x, u) = \begin{pmatrix} \frac{\partial \phi}{\partial x_1}(x, u) \\ \vdots \\ \frac{\partial \phi}{\partial x_n}(x, u) \end{pmatrix}$ o gradiente

de ϕ em relação à variável x e por $\nabla_u \phi(x, u) = \begin{pmatrix} \frac{\partial \phi}{\partial u_1}(x, u) \\ \vdots \\ \frac{\partial \phi}{\partial u_n}(x, u) \end{pmatrix}$ o gradiente de ϕ em relação à variável u . Digamos (x^*, u^*) é um ponto de sela de ϕ , isto é, $\phi(x, u^*) \leq \phi(x^*, u^*) \leq \phi(x^*, u)$, para $\forall x \geq 0, \forall u \geq 0$.

Lema 2: Seja uma função $\phi: \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow \mathbb{R}$ e $(x^*, u^*) \geq 0$ é um ponto de sela de ϕ , então

i) $\frac{\partial \phi}{\partial x_i}(x^*, u^*) \leq 0, i = 1, \dots, n$; $\nabla_x \phi(x^*, u^*) \cdot x^* = 0$;

ii) $\frac{\partial \phi}{\partial u_i}(x^*, u^*) \geq 0, i = 1, \dots, n$; $\nabla_u \phi(x^*, u^*) \cdot u^* = 0$.

Prova: fixado o vetor u^* como $\phi(x, u^*) \leq \phi(x^*, u^*)$, então cada $x_i^* \geq 0$ é máximo da função $\phi(x, u)$, em que apenas $x_i \geq 0$ está variando, logo, pelo Lema 1, $\frac{\partial \phi}{\partial x_i}(x^*, u^*) \leq 0$ e $\frac{\partial \phi}{\partial x_i}(x^*, u^*) \cdot x_i^* = 0$. Idem para o item ii).

Lema 3: Seja uma função $\phi: \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow \mathbb{R}$, tal que

i) $\frac{\partial \phi}{\partial x_i}(x^*, u^*) \leq 0, i = 1, \dots, n$; $\nabla_x \phi(x^*, u^*) \cdot x^* = 0$;

$$\text{ii)} \quad \frac{\partial \phi}{\partial u_i}(x^*, u^*) \geq 0, i=1, \dots, n; \quad \nabla_u \phi(x^*, u^*) \cdot u_i^* = 0;$$

$$\text{iii)} \quad \phi(x, u^*) \leq \phi(x^*, u^*) + \nabla_x \phi(x^*, u^*) \cdot (x - x^*), \text{ isto é, } \phi(x, u^*) \text{ é côncava na variável } x;$$

$$\text{iv)} \quad \phi(x^*, u) \geq \phi(x^*, u^*) + \nabla_u \phi(x^*, u^*) \cdot (u - u^*), \text{ isto é, } \phi(x^*, u) \text{ é convexa na variável } u \text{ então } (x^*, u^*) \geq 0 \text{ é um ponto de sela de } \phi.$$

Prova: Como $\phi(x, u^*) \leq \phi(x^*, u^*) + \nabla_x \phi(x^*, u^*) \cdot (x - x^*)$ e $\nabla_x \phi(x^*, u^*) \cdot x^* = 0$, então $\phi(x, u^*) \leq \phi(x^*, u^*) + \nabla_x \phi(x^*, u^*) \cdot x \leq \phi(x^*, u^*)$, pois $\frac{\partial \phi}{\partial x_i}(x^*, u^*) \leq 0, i=1, \dots, n$ e, portanto, $\nabla_x \phi(x^*, u^*) \cdot x \leq 0$. Logo, $\phi(x, u^*) \leq \phi(x^*, u^*)$.

Como $\phi(x^*, u) \geq \phi(x^*, u^*) + \nabla_u \phi(x^*, u^*) \cdot (u - u^*)$ e $\nabla_u \phi(x^*, u^*) \cdot u^* = 0$, então $\phi(x^*, u) \geq \phi(x^*, u^*) + \nabla_u \phi(x^*, u^*) \cdot u \geq \phi(x^*, u^*)$, pois $\frac{\partial \phi}{\partial u_i}(x^*, u^*) \geq 0, i=1, \dots, n$ e, portanto, $\nabla_u \phi(x^*, u^*) \cdot u \geq 0$. Logo, $\phi(x^*, u) \geq \phi(x^*, u^*)$. Logo, $\phi(x, u^*) \leq \phi(x^*, u^*) \leq \phi(x^*, u)$.

Consideremos o problema:

$$\begin{aligned} \max \quad & g(x_1, \dots, x_n) \\ \text{sa} \quad & f_1(x_1, \dots, x_n) \geq 0; \\ & \vdots \\ & f_m(x_1, \dots, x_n) \geq 0; \\ & x_1 \geq 0, \dots, x_n \geq 0. \end{aligned}$$

Definamos o lagrangeano $\phi: \mathbb{R}_+^n \times \mathbb{R}_+^m \rightarrow \mathbb{R}$ do problema acima como

$$\phi(x, u) = g(x) + u_1 \cdot f_1(x) + \dots + u_m \cdot f_m(x).$$

Digamos que (x^*, u^*) é uma sela do lagrangeano acima, isto é, $\phi(x, u^*) \leq \phi(x^*, u^*) \leq \phi(x^*, u)$. Não é difícil mostrar que x^* é um máximo global para g .

Primeiramente, observemos que x^* está na região viável, pois $\phi(x^*, u^*) \leq \phi(x^*, u)$, isto é, $g(x^*) + u_1^* \cdot f_1(x^*) + \dots + u_m^* \cdot f_m(x^*) \leq g(x^*) + u_1 \cdot f_1(x^*) + \dots + u_m \cdot f_m(x^*)$. Fazendo, $u_i = (u_i^* + 1)$ e $u_k = u_k^*, k \neq i$, obtemos na expressão acima $f_i(x^*) \geq 0$.

Em segundo lugar, observe que se fazemos $u_i = 0, i=1, \dots, m$, obtemos na mesma expressão $u_1^* \cdot f_1(x^*) + \dots + u_m^* \cdot f_m(x^*) \leq 0$; por outro lado, se $u_i = 2 \cdot u_i^*, i=1, \dots, m$, obtemos $u_1^* \cdot f_1(x^*) + \dots + u_m^* \cdot f_m(x^*) \geq 0$. Logo, $\sum_{i=1}^m u_i^* \cdot f_i(x^*) = 0$.

Em terceiro lugar, como $\phi(x, u^*) \leq \phi(x^*, u^*)$, isto é, $g(x) + u_1^* \cdot f_1(x^*) + \dots + u_m^* \cdot f_m(x^*) \leq g(x) + u_1^* \cdot f_1(x^*) + \dots + u_m^* \cdot f_m(x^*)$, como $\sum_{i=1}^m u_i^* \cdot f_i(x^*) = 0$, então $g(x) \leq g(x^*)$ e x^* é

máximo global para g .

Observe que $\sum_{i=1}^{l+m} u_i^* \cdot f_i(x^*) = 0$ implica que $u_i^* \cdot f_i(x^*) = 0, i=1, \dots, m$, que é a segunda condição de Kuhn-Tucker se $u_1^*, \dots, u_m^*$ são os multiplicadores; pela forma como o lagrangeano foi definido, e porque está sendo assegurada a existência $u^* \geq 0$, então também a terceira condição de Kuhn-Tucker é válida, isto é, $u_1^*, \dots, u_m^* \geq 0$.

Digamos que x^* é um máximo local para $g(x)$ na região determinada pelas restrições. Digamos que as fronteiras $f_1(x^*)=0, \dots, f_k(x^*)=0$ são ativas e que as demais são inativas, isto é, $f_{k+1}(x^*) > 0, \dots, f_m(x^*) > 0$.

Seja $x: [0, +\infty) \rightarrow \Omega$ uma **trajetória admissível** em x^* , ou seja, $x(0) = x^*$ e

$$\begin{aligned} \left. \frac{d}{dt} f_1(x(t)) \right|_{t=0} &= \nabla f_1(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla f_1(x^*) \cdot x'(0) \geq 0, \\ &\vdots \\ \left. \frac{d}{dt} f_k(x(t)) \right|_{t=0} &= \nabla f_k(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla f_k(x^*) \cdot x'(0) \geq 0. \end{aligned}$$

Portanto,

$$(x'_1(0) \ \dots \ x'_n(0)) \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x^*) & \dots & \frac{\partial f_m}{\partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n}(x^*) & \dots & \frac{\partial f_m}{\partial x_n}(x^*) \end{pmatrix} = (\nabla f_1(x^*) \cdot x'(0) \ \dots \ \nabla f_m(x^*) \cdot x'(0)) \geq 0$$

Como x^* é um máximo local para g , então $\left. \frac{d}{dt} g(x(t)) \right|_{t=0} = \nabla g(x(t)) \cdot x'(t) \Big|_{t=0} = \nabla g(x^*) \cdot x'(0) \leq 0$, isto é, $-\nabla g(x^*) \cdot x'(0) \geq 0$.

Pelo **Lema de Farkas**, fazemos $x'(0) = y$,

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x^*) & \dots & \frac{\partial f_m}{\partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n}(x^*) & \dots & \frac{\partial f_m}{\partial x_n}(x^*) \end{pmatrix} = A$$

e $-\nabla g(x^*) = b$, então obtemos que $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$. Lembremos que para que possamos aplicar o Lema de Farkas, temos que verificar a proposição $y' \cdot A \geq 0 \rightarrow y' \cdot b \geq 0$ para todo $y \in \mathbb{R}^n$, tal que $y' \cdot A \geq 0$, logo, temos que assumir a hipótese de qualificação de restrição: para todo $y \in \mathbb{R}^n$, tal que $y' \cdot A \geq 0$,

existe uma **trajetória admissível** $x(t)$ em x^* , tal que $x'(0)=y$. Em outras palavras, $\forall y \in \mathbb{R}^n, \nabla f_1(x^*) \cdot y \geq 0, \dots, \nabla f_k(x^*) \cdot y \geq 0, \exists x(t), x(0)=x^* \rightarrow x'(0)=y$.

Admitida a hipótese acima pelo Lema de Farkas, existe $\begin{pmatrix} u_1^* \\ \vdots \\ u_m^* \end{pmatrix} \geq 0$, tal que

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x^*) & \dots & \frac{\partial f_m}{\partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n}(x^*) & \dots & \frac{\partial f_m}{\partial x_n}(x^*) \end{pmatrix} \begin{pmatrix} u_1^* \\ \vdots \\ u_m^* \end{pmatrix} = \begin{pmatrix} -\frac{\partial g}{\partial x_1}(x^*) \\ \vdots \\ -\frac{\partial g}{\partial x_n}(x^*) \end{pmatrix},$$

Isto é, $u_1^* \cdot \nabla_{x_1} f_1(x^*) + \dots + u_m^* \cdot \nabla_{x_m} f_m(x^*) = -\nabla_x g(x^*)$, ainda, $\nabla_x g(x^*) + u_1^* \cdot \nabla_{x_1} f_1(x^*) + \dots + u_m^* \cdot \nabla_{x_m} f_m(x^*) = 0$.

Seja $\phi(x, u) = g(x) + u_1 \cdot f_1(x) + \dots + u_m \cdot f_m(x)$, então $\nabla_x \phi(x, u) = \nabla_x g(x^*) + u_1^* \cdot \nabla_{x_1} f_1(x^*) + \dots + u_m^* \cdot \nabla_{x_m} f_m(x^*) = 0$ e $\nabla_u \phi(x^*, u^*) = \begin{pmatrix} f_1(x^*) \\ \vdots \\ f_m(x^*) \end{pmatrix} \geq 0$ e $\nabla_u \phi(x^*, u^*) \cdot u^* = 0$.

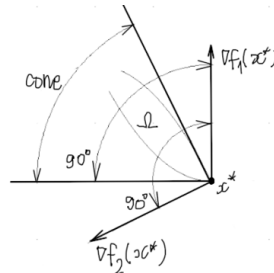
A convexidade de $\phi(x, u) = g(x) + u_1 \cdot f_1(x) + \dots + u_m \cdot f_m(x)$ na variável $u = (u_1, \dots, u_m)$ resulta do fato que $\phi(x, u)$ é linear em u . Se $g(x), f_1(x), \dots, f_m(x)$ são côncavas, como $u_1 \geq 0, \dots, u_m \geq 0$, então $\phi(x, u) = g(x) + u_1 \cdot f_1(x) + \dots + u_m \cdot f_m(x)$ é côncava. Portanto, (x^*, u^*) é uma sela de $\phi(x, u)$.

8. As qualificações de restrições

A razão pela qual tivemos que pospor os comentários sobre a hipótese de “qualificação de restrição” para esta lição é que as hipóteses alternativas envolvem as propriedades de convexidade das funções implicadas.

No artigo “*Lagrange Multipliers Revisited*”, Morton Slater aprimorou o artigo de Kuhn e Tucker contornando o uso que os autores fizeram da diferenciabilidade das funções e mostrando que a bastava a hipótese da concavidade. Se olharmos a seção anterior, observamos como fortemente a diferenciabilidade foi utilizada. Um subproduto do trabalho de Slater é que a “restrição de qualificação” de Kuhn e Tucker pode ser substituída por uma condição mais direta, chamada “Condição de Slater”, expressa na segunda

condição do Teorema 1²¹: caso as restrições $f_h(x) \geq 0$ sejam tais que as funções côncavas $f_h(x)$, $h=1, \dots, m$, sejam côncavas e existe $\bar{x}$, tal que $f_h(\bar{x}) > 0$, $h=1, \dots, m$, então obtemos o Teorema de Kuhn-Tucker. Em seu artigo²² já referido, David Peterson mostra que a condição de Slater implica a condição de qualificação de Kuhn-Tucker. O desenho abaixo heurísticamente nos convence disso.



A região Ω é determinada por $f_1(x) \geq 0$ e $f_2(x) \geq 0$. Se $f_1(x)$ e $f_2(x)$ são côncavas, então os contornos superiores $f_1(x) \geq 0$ e $f_2(x) \geq 0$ são conjuntos convexos e, portanto, a interseção desse contorno é um conjunto convexo. O desenho é tal que existem $\bar{x} \in \Omega$, $f_1(\bar{x}) > 0$ e $f_2(\bar{x}) > 0$. A relação dos contornos superiores com os gradientes em x^* é tal que Ω está dentro do cone $C_{\pi/2}$ determinado pelos gradientes $\nabla f_1(x^*)$ e $\nabla f_2(x^*)$. Portanto, a condição de qualificação de Kuhn-Tucker é satisfeita.

Na observação acima, a hipótese importante é que os contornos superiores $f_1(x) \geq 0$ e $f_2(x) \geq 0$ são conjuntos convexos, ou seja, que as funções $f_1(x)$ e $f_2(x)$ sejam quase-côncavas. Isto é exatamente o que fazem Arrow e seus coautores no corolário 4 do artigo "*Constraint Qualifications in Maximization Problems*", de 1961²³, e no corolário 3 do "*Constraint Qualifications in Maximization Problems II*", de 1960²⁴.

Outra condição que implica a condição de qualificação de Kuhn-Tucker²⁵ é se no ponto crítico do lagrangeano $x^* \in \Omega$, os gradientes das fronteiras ativas geram um espaço de dimensão igual ao número de fronteiras ativas. Esses resultados estão, respectivamente, nos corolários 6 e 4 dos artigos acima.

Como vimos no exemplo 12 da Lição 5, não é razoável exigirmos que os gradientes das restrições ativas sejam necessariamente linearmente independentes nos pontos críticos do lagrangeano. Na Lição 4, impusemos essa condição com o objetivo de que os gradientes $\nabla f_i(\bar{x}(t))$, $1 \leq i \leq k$ definissem uma superfície s de dimensão $n-k$, pois, nesse caso, o subespaço gerado pelos vetores tangentes $\bar{x}(t)$, ou seja, o “plano” tangente em $x(t)$, tem dimensão $n-k$. Formulamos essa condição algebricamente exigindo que o posto da matriz

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_k}{\partial x_1}(x) & \frac{\partial f_k}{\partial x_2}(x) & \dots & \frac{\partial f_k}{\partial x_n}(x) \end{pmatrix}$$

seja k . O importante aqui é que o plano tangente

tenha dimensão $n-k$. Se o posto da matriz acima é k , então os vetores $\nabla f_i(\bar{x}(t))$, $1 \leq i \leq k$ são linearmente independentes. Não consideramos situações em que há mais que k restrições. Se tivermos $m > k$ restrições

e exigirmos que o posto da matriz $\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \dots & \dots & \dots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$ seja k , então o plano

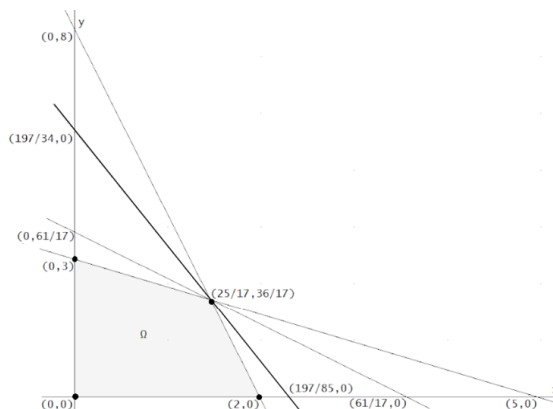
tangente tem dimensão $n-k$ e podemos retirar do conjunto de gradientes $\nabla f_i(\bar{x}(t))$, $1 \leq i \leq m$, k gradientes linearmente independentes. Portanto, o importante é exigir que o posto da matriz seja k , e foi por simplicidade que na Lição 4 não apresentamos mais que k restrições.

Exemplo 16: Seja o problema:

$$\begin{aligned} \max \quad & 5 \cdot x + 2 \cdot y \\ \text{sa} \quad & 3 \cdot x + 5 \cdot y \leq 15 \\ & 4 \cdot x + y \leq 8 \\ & x + y \leq \frac{61}{17} \\ & x \geq 0, y \geq 0. \end{aligned}$$

O máximo está no ponto $(\frac{25}{17}, \frac{36}{17})$. Nesse ponto, as fronteiras $3 \cdot x + 5 \cdot y = 15$, $4 \cdot x + y = 8$ e $x + y = \frac{61}{17}$ são ativas. Os três gradientes $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$, $\begin{pmatrix} 4 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ são obviamente linearmente dependentes: $\text{posto} \begin{pmatrix} 3 & 5 \\ 4 & 1 \\ 1 & 1 \end{pmatrix} = 2$. A restrição $x + y = \frac{61}{17}$ é redundante

para a determinação da região Ω e pode ser dispensada. Observe que $\begin{pmatrix} 5 \\ 2 \end{pmatrix} = \frac{3}{17} \cdot \begin{pmatrix} 3 \\ 5 \end{pmatrix} + \frac{19}{17} \cdot \begin{pmatrix} 4 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, ou seja, as condições de Kuhn-Tucker são satisfeitas no ponto $\left(\frac{25}{17}, \frac{36}{17}\right)$.



Num problema no plano $\mathbb{R}^2$, qualquer vértice da região Ω sempre está definido por duas restrições, cujos gradientes são linearmente independentes, como podemos verificar no exemplo acima e também em todos os exemplos que já vimos. Outros pontos na fronteira da região Ω que não são vértices, estão sobre arestas, são determinados por uma restrição. Se a região Ω está contida no $\mathbb{R}^3$, três restrições determinam um vértice, duas restrições determinam uma aresta e uma restrição determina uma face do poliedro.

Exemplo 17: No problema primal
$$\begin{cases} \max & b' \cdot y \\ \text{s.t.} & A \cdot y \leq a, \\ & y \geq 0 \end{cases}$$
 a restrição $A \cdot y \leq a, y \geq 0$, em que A é uma matriz $m \times n$, qualquer vértice é sempre determinado por n restrições, tais que nesses vértices os gradientes dessas restrições são linearmente independentes.

Se o ponto que satisfaz as condições de Kuhn-Tucker para um máximo é um vértice da região Ω , então esse vértice é determinado por n restrições, tais que os n vetores $\begin{pmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{pmatrix}$ são linearmente independentes, e isso implica que os correspondentes multiplicadores $x_i \geq 0$ são unicamente determinados. E assim, podemos

observar em relação aos demais pontos, ou seja, pontos sobre arestas ou faces. Para cada ponto que satisfaz as condições de Kuhn-Tucker para um máximo no problema primal, existe apenas um conjunto de multiplicadores $x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$, $A' \cdot x \geq b$.

Nos exemplos 11 e 12 da Lição 5, no vértice $(1,0)$ da região Ω , os dois gradientes das restrições são linearmente dependentes e não geram o $\square^2$. Chamamos um vértice como esse de uma “cúspide” e, portanto, o Teorema de Kuhn-Tucker é válido apenas em alguns casos particulares para pontos de mínimos ou máximos nesse ponto.

O resultado de Arrow e seus coautores está afirmando que a “dimensão” do ponto crítico, ou seja, a de fronteiras ativas nesse ponto, geram um espaço de dimensão igual ao seu número, então a qualificação de restrição é válida. Veja que no exemplo 15, temos no vértice $\left(\frac{25}{17}, \frac{26}{17}\right)$ três fronteiras ativas gerando um espaço de dimensão dois, e vale a restrição de qualificação, o que aparentemente é uma contradição em relação ao resultado de Arrow. Mas, não é! A fronteira $x+y=\frac{61}{17}$ é supérflua na determinação da região Ω . Não é o caso no exemplo da cúspide.

Como consequência desse resultado, a qualificação de restrição é sempre satisfeita em problemas lineares.

¹ BONNESEN, T.; FENCHEL, W. **Theory of Convex Bodies**. Tradução do original em alemão de 1933. USA: BSC Associates, 1987, p. viii.

² KJELDSSEN, TINNE H., “Egg-Forms and Measure-Bodies: Different Mathematical Practices in the Early History of the Modern Theory of Convexity”. **Science in Context**, v. 22, nº 1 (2009), p. 89.

³ ROBERTS, A. W.; VARBERG, D. E. **Convex Functions**. New York: Academic Press, 1973, p. 8.

⁴ GUERRAGGIO, A.; MOLHO, E. “The Origins of Quasi-Concavity: a Development between Mathematics and Economics”. **Historia Mathematica**, v. 31 (2004), p. 63.

- ⁵ FENCHEL, W. **Convex Sets, Cones and Functions**. Princeton University, 1956.
- ⁶ KLEE, V. "What is a Convex Set?". **The American Mathematical Monthly**, v. 78, nº 6 (1971), p. 624.
- ⁷ KUHN, H. W.; TUCKER, A. W. (Ed.). **Linear Inequalities and related Systems**. USA: Princeton University Press, 1956.
- ⁸ NEUMANN, J. "A Model of General Economic Equilibrium". **The Review of Economic Studies**, v. 13, nº 1 (1945-6), p. 1.
- ⁹ DEBREU, G. **Theory of Value An Axiomatic Analysis of Economic Equilibrium**. Yale: Cowles Foundations, Monograph 17, 1959, p.x
- ¹⁰ MCKENZIE, L. "On Equilibrium in Graham's Model of World Trade and Other Competitive Systems". **Econometrica**, v. 22, nº 2 (1954), pp. 147-161.
- ¹¹ ARROW, K.; DEBREU, G. "Existence of an Equilibrium for a Competitive Economy". **Econometrica**, v. 22, nº 3 (1954), pp. 147-161.
- ¹² STARR, R. **General Equilibrium Theory An Introduction**. UK: Cambridge University Press, 1997, pp. 77, 91.
- ¹³ NASH, J. "Non-Cooperative Games." **Doctoral Dissertation**, Princeton University, 1950.
- ¹⁴ NASH, J. "Non-cooperative Games". **The Annals of Mathematics**, Second Series, v. 54, nº 2 (1951), pp. 286-295.
- ¹⁵ PARK, S. "Ninety Years of the Brouwer Fixed Theorem". **Vietnam Journal of Mathematics**, v. 27, nº 3 (1999), pp. 187-222.
- ¹⁶ GUERRAGGIO, A.; MOLHO, E. "The Origins of Quasi-Concavity: a Development between Mathematics and Economics". **Historia Mathematica**, v. 31 (2004), p. 64.
- ¹⁷ NEUMANN, J. "On the Theory of Games of Strategy". In: TUCKER, A.; LUCE, R. (Ed.). **Contributions to the Theory of Games, v. 4**. Annals of Mathematics Studies, nº 40. Ed. Princeton: Princeton University Press, 1959, p. 26.
- ¹⁸ GUERRAGGIO, A.; MOLHO, E. "The Origins of Quasi-Concavity: a Development between Mathematics and Economics". **Historia Mathematica**, v. 31 (2004), p. 69.
- ¹⁹ SIMONSEN, Mário H. **Teoria Microeconômica**, v. I. Rio de Janeiro: Editora da FGV, 1985, p. 40.
- ²⁰ KUHN, H.; TUCKER, A. "Nonlinear Programming". In: NEYMAN, Jerzy (Ed.). **Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1950, p. 481.
- ²¹ SLATER, M. "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming". **Cowles Foundation Discussion Paper**, nº 80 (1950), p. 2.
- ²² PETERSON, D. "A Review of Constraint Qualifications in Finite-Dimensional Spaces". **SIAM Review**, v. 15, nº 3 (1973), p. 645.
- ²³ ARROW, K.; HURWICZ, L.; UZAWA, H. "Constraint Qualifications in Maximization Problems". **Naval Research Logistics**, v. 8, nº 2 (1961).

²⁴ ARROW, K.; UZAWA, H. "Constraint Qualifications in Maximization Problems II". **Technical Report n. 84**, Institute for Mathematical Studies in Social Sciences, Stanford University, 1960.

²⁵ PETERSON, D. "A Review of Constraint Qualifications in Finite-Dimensional Spaces". **SIAM Review**, v. 15, nº 3 (1973), p. 645.

Lição SETE

Condições de 2ª Ordem

Sumário

Lição 7 – Condições de 2ª Ordem

- 1.Introdução
2. Formas quadráticas e suas definições
3. Redução da forma quadrática
4. A derivada segunda
5. A determinação da definição da forma quadrática
 - 5.1 Menores principais
 - 5.2 Critério de Jacobi
6. Formas quadráticas restritas a subespaços
 - 6.1. A demonstração de Väliaho para o Teorema de Mann-Debreu
7. Teorema de Arrow-Enthoven
- 8.Critérios para Mínimos e Máximos Locais com Restrições, Desigualdades e Igualdades
9. Notas históricas
 - 9.1 Teorema de Hancock
 - 9.2 Hotelling e as utilidades
 - 9.3 A demonstração de Mann
 - 9.4 O apêndice matemático de Samuelson
 - 9.5 A demonstração de Debreu
 - 9.6 O teorema de Arrow-Enthoven

Lição 7

Condições de 2ª Ordem

1.Introdução

Nosso objetivo nesta lição é analisar a posição relativa entre o plano tangente e o gráfico de f na vizinhança de um ponto $x^* \in \mathbb{R}^n$, no qual $f = y \in \mathbb{R}$ está definida. Essa relação é essencial para a determinação das condições de 2ª ordem, isto é, aquelas condições que asseguram que os pontos que satisfazem as condições de 1ª ordem para mínimo ou máximo são efetivamente mínimos ou máximos.

Como vimos, as definições de funções convexas e côncavas se traduzem em termos da relação entre os gráficos das funções e os planos que os tangenciam: se $f(x) \leq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, então f é côncava; se $f(x) \geq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$, então f é convexa. As relações anteriores determinam se um ponto crítico x^* de f , isto é, $\nabla f(x^*) = 0$, é um mínimo ou um máximo, caso f seja convexa ou côncava, respectivamente. Portanto, a análise que empreendemos nesta lição tem por objetivo estender essa ideia simples para funções em geral.

Para a determinação das relações entre gráficos e planos, estudamos um tipo particular de funções que são chamadas de formas quadráticas.

As derivadas segundas $\nabla^2 f$ de funções de classe C^2 , ou seja, que possuem derivadas parciais de segunda ordem contínuas, são formas quadráticas. Fixado um ponto x^* , $\nabla^2 f(x^*)$ determina a posição em

relação ao eixo y do plano tangente $y = f(x^*) + \nabla f(x^*) \cdot (x - x^*)$ e do gráfico $(x, f(x))$ na vizinhança de x^* . A posição relativa é indicada pelo que se diz a “definição da forma quadrática”, ou seja, se $\nabla^2 f(x^*)(x_1, \dots, x_n)$ é positiva, negativa, não-negativa ou não-positiva, ou se assume valores positivos e negativos em $\mathbb{R}^n \setminus \{0\}$.

Os critérios que possibilitam determinar a definição de uma forma quadrática em $\mathbb{R}^n$ ou num subespaço de $\mathbb{R}^n$ surgiram pouco a pouco ao longo de 200 anos, começando na segunda metade do século 18 e se estendendo até a primeira metade do século 20. Em 1952, Gerard Debreu (1921-2004) reuniu esses resultados e providenciou demonstrações adequadas num artigo que atualmente serve de referência¹. As referências indicadas por Debreu dão uma noção do período através do qual a análise se estendeu. A referência mais antiga é Lagrange, em 1758, e a mais recente é o artigo “*Quadratic Forms with Linear Constraints*”² do austríaco Henry Berthold Mann (1905-2000).

2. Formas quadráticas e suas definições

Uma forma quadrática $Q(x_1, \dots, x_n)$ em $\mathbb{R}^n$ é um polinômio homogêneo de grau dois (por isso o adjetivo “quadrática”) definido nas variáveis $x_1, \dots, x_n$:

$$Q(x_1, \dots, x_n) = \sum_{1 \leq i, j \leq n} a_{ij} x_i \cdot x_j.$$

Podemos representar Q como $x' \cdot A \cdot x$, em que $A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ é uma matriz simétrica, isto é, $a_{ij} = a_{ji}$, $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ e $x' = (x_1, \dots, x_n)$ são o transposto do vetor x . Observe que:

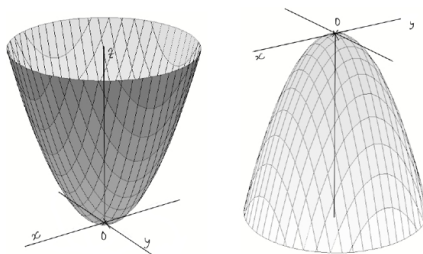
$$Q(x) = x' \cdot A \cdot x = \sum_{1 \leq i \leq n} a_{ii} \cdot x_i^2 + 2 \cdot \sum_{1 \leq i < j \leq n} a_{ij} \cdot x_i \cdot x_j.$$

São exemplos de formas quadráticas:

$Q(x_1) = -5 \cdot x_1^2$; $Q(x_1, x_2) = 6 \cdot x_1^2 + 2 \cdot x_1 \cdot x_2 - 5 \cdot x_2^2 = (x_1 \ x_2) \cdot \begin{pmatrix} 6 & 1 \\ 1 & -5 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$; $Q(x_1, x_2, x_3) = 4 \cdot x_1^2 + x_2^2 - 3 \cdot x_3^2 + 5 \cdot x_1 \cdot x_2 + 6 \cdot x_1 \cdot x_3 - 6 \cdot x_2 \cdot x_3$, isto é,

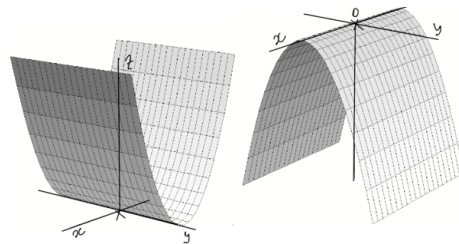
$$Q(x_1, x_2, x_3) = (x_1 \ x_2 \ x_3) \cdot \begin{pmatrix} 4 & \frac{5}{2} & 3 \\ \frac{5}{2} & 1 & -3 \\ 3 & -3 & -3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Elucidemos o que chamamos a “definição da forma”. Dizemos que $Q(x)$ é positiva-definida se $Q(x) > 0$, para todo $0 \neq x \in \mathbb{R}^n$; negativa-definida se $Q(x) < 0$, para todo $0 \neq x \in \mathbb{R}^n$. Por exemplo, a forma $z(x, y) = (x \ y) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2$ é positiva-definida e $z(x, y) = (x \ y) \cdot \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = -x^2 - y^2$ é negativa-definida no $\mathbb{R}^2$. Os gráficos dessas funções no $\mathbb{R}^3$ são chamados de paraboloides elípticos e estão abaixo:



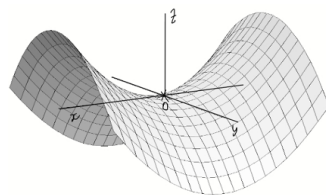
Nos dois exemplos, observe que as isoquantas de $z(x, y)$ são circunferências $x^2 + y^2 = r^2$, que definem apenas um ponto como $x^2 + y^2 = 0$ ou não definem nenhum ponto no plano, como no caso $x^2 + y^2 = -1$.

A forma é dita positiva-semidefinida se $Q(x) \neq 0$ e $Q(x) \geq 0$, para todo $x \in \mathbb{R}^n$; negativa-semidefinida se $Q(x) \neq 0$ e $Q(x) \leq 0$, para todo $x \in \mathbb{R}^n$. As formas $z(x, y) = (x \ y) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = x^2$ e $z(x, y) = (x \ y) \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = y^2$ são positivas-semidefinidas no $\mathbb{R}^2$; as formas $z(x, y) = (x \ y) \cdot \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = -x^2$ ou $z(x, y) = (x \ y) \cdot \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = -y^2$ são negativas-semidefinidas. Os gráficos são chamados de cilindros hiperbólicos. Abaixo estão os gráficos de $z(x, y) = x^2$ e $z(x, y) = -y^2$:

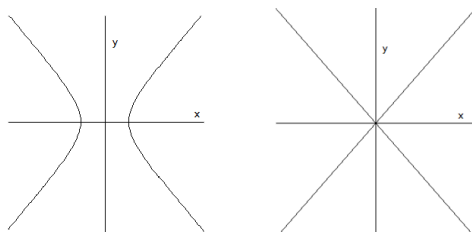


Observe que as isoquantas de $z(x, y)$, no caso dos cilindros parabólicos, são duas retas paralelas, $x = \pm\sqrt{c}$, como $x^2 = c$, uma reta, $x = 0$, como $x^2 = 0$, ou não definem nenhum ponto no plano como no caso $x^2 = -1$.

A forma é dita indefinida ou não-definida se existem $x_1 \in \mathbb{R}^n$ e $x_2 \in \mathbb{R}^n$, tal que $Q(x_1) < 0$ e $Q(x_2) > 0$. As formas $z(x, y) = (x \ y) \cdot \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = -x^2 + y^2$ e $z(x, y) = (x \ y) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = x^2 - y^2$ são não-definidas. Os gráficos são chamados de paraboloides hiperbólicos ou selas. Abaixo, $z(x, y) = x^2 - y^2$:



Observe que as isoquantas de $z(x, y)$, no caso dos cilindros hiperbólicos, são hipérboles, como $x^2 - y^2 = 1$, ou duas retas concorrentes como $x^2 - y^2 = 0$. Veja os gráficos abaixo.



Para formas quadráticas definidas no plano, as formas anteriores $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ e $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ são todas as combinações possíveis de $-1, 0$ e 1 . Através de trocas de base, podemos reduzir qualquer forma definida no plano a alguma das expressões anteriores.

Em resumo, considerando a forma $Q(x) = x' \cdot A \cdot x$, dizemos que $Q(x)$ é positiva-definida se $Q(x) > 0$, para todo $0 \neq x \in \mathbb{R}^n$; negativa-definida se $Q(x) < 0$, para todo $0 \neq x \in \mathbb{R}^n$; positiva-semidefinida se $Q(x) \geq 0$, para todo $x \in \mathbb{R}^n$; negativa-semidefinida se $Q(x) \leq 0$, para todo $x \in \mathbb{R}^n$; indefinida ou não-definida se existem $x_1 \in \mathbb{R}^n$ e $x_2 \in \mathbb{R}^n$, tais que $Q(x_1) < 0$ e $Q(x_2) > 0$. Dada uma forma, nos referimos ao caso particular que ela satisfaz como a “definição da forma”.

3. Redução da forma quadrática

Para expressar as formas quadráticas da maneira que queremos, utilizamos trocas de coordenadas, substituindo na forma quadrática as variáveis $x_1, \dots, x_n$ pelas variáveis $\bar{x}_1, \dots, \bar{x}_n$. Para isso, utilizamos a passagem do sistema de coordenadas $x_1, \dots, x_n$ para o sistema $\bar{x}_1, \dots, \bar{x}_n$ dado por:

$$\begin{aligned}\bar{x}_1 &= l_{11} \cdot x_1 + l_{12} \cdot x_2 + \dots + l_{1n} \cdot x_n, \\ \bar{x}_2 &= l_{21} \cdot x_1 + l_{22} \cdot x_2 + \dots + l_{2n} \cdot x_n, \\ &\dots \\ \bar{x}_n &= l_{n1} \cdot x_1 + l_{n2} \cdot x_2 + \dots + l_{nn} \cdot x_n.\end{aligned}$$

A troca das variáveis $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ para $\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{pmatrix}$ é expressa por $\begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{pmatrix} = \bar{x} = L \cdot x = \begin{pmatrix} l_{11} & \dots & l_{1n} \\ \vdots & \ddots & \vdots \\ l_{n1} & \dots & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, em que $L = \begin{pmatrix} l_{11} & \dots & l_{1n} \\ \vdots & \ddots & \vdots \\ l_{n1} & \dots & l_{nn} \end{pmatrix}$ é dita a matriz de troca de coordenadas de $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ para $\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{pmatrix}$. A única exigência é $\det L \neq 0$.

Digamos que $Q(\bar{x}) = \bar{x}' \cdot \bar{A} \cdot \bar{x}$. Utilizando a matriz da troca de coordenadas do sistema x para $\bar{x}$, $\bar{x} = L \cdot x$, então $Q(\bar{x}) = \bar{x}' \cdot \bar{A} \cdot \bar{x} = (L \cdot x)' \cdot \bar{A} \cdot (L \cdot x) = x' \cdot (L' \cdot \bar{A} \cdot L) \cdot x = Q(x)$. Observe que $(L' \cdot \bar{A} \cdot L)' = L' \cdot \bar{A}' \cdot L = L' \cdot \bar{A} \cdot L$, portanto $L' \cdot \bar{A} \cdot L = A$ é a matriz simétrica da forma em relação ao sistema de coordenadas $x_1, \dots, x_n$. Repetindo, $L' \cdot \bar{A} \cdot L$ é a matriz da forma quadrática no sistema de coordenadas $x_1, \dots, x_n$; a matriz L é a matriz da troca de coordenadas de $x_1, \dots, x_n$ para $\bar{x}_1, \dots, \bar{x}_n$; A matriz $(L^{-1})' \cdot A \cdot (L^{-1}) = \bar{A}$ é matriz associada à forma no sistema de coordenadas $\bar{x}$, em que L^{-1} é a troca de base de $\bar{x}_1, \dots, \bar{x}_n$ para $x_1, \dots, x_n$. $Q(x)$ é um número real que não é

alterado pela troca de coordenadas:

$$\mathcal{Q}(x) = x' \cdot A \cdot x = (L^{-1} \cdot \bar{x})' \cdot \bar{A} \cdot (L^{-1} \cdot \bar{x}) = \bar{x}' \cdot (L^{-1})' \cdot A' \cdot (L^{-1}) \cdot \bar{x} = \bar{x}' \cdot \bar{A} \cdot \bar{x} = \mathcal{Q}(\bar{x}).$$

Reduzir a forma quadrática a uma soma de quadrados é encontrar um sistema de coordenadas $\bar{x}$, tal que a matriz $\bar{A} = K' \cdot A \cdot K$ seja uma matriz diagonal, isto é,

$$\bar{A} = \begin{pmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & c_n \end{pmatrix}.$$

A matriz K é a troca de base de $\bar{x}_1, \dots, \bar{x}_n$ para $x_1, \dots, x_n$. No sistema $\bar{x}$, portanto, a expressão de $\mathcal{Q}$ é $\mathcal{Q}_1(\bar{x}_1, \dots, \bar{x}_n) = c_1 \cdot (\bar{x}_1)^2 + \dots + c_n \cdot (\bar{x}_n)^2$. Nesse caso, a forma foi reduzida a uma soma de quadrados.

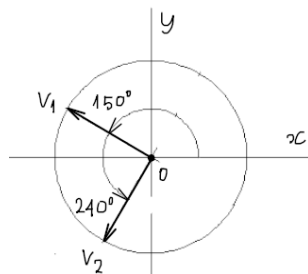
Nos cursos de Álgebra Linear, aprendemos que uma matriz simétrica é diagonalizável, ou seja, existe uma troca de base dada por uma matriz ortogonal P , isto é, $P' \cdot P = I$, tal que $P' \cdot A \cdot P = \bar{A}$, uma matriz diagonal. Mais adiante neste curso, voltaremos a esse resultado.

Exemplo 1: Seja a matriz simétrica $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. O polinômio característico de A é $p(\lambda) = \lambda^2 - (a+c) \cdot \lambda + a \cdot c - b^2$, cujas raízes são reais $\lambda_1 = \frac{(a+c) + \sqrt{(a-c)^2 + 4 \cdot b^2}}{2}$ e $\lambda_2 = \frac{(a+c) - \sqrt{(a-c)^2 + 4 \cdot b^2}}{2}$. Obviamente, $\lambda_1 \geq \lambda_2$. Digamos que $\lambda_1 \neq \lambda_2$ e que v_1 e v_2 são autovetores associados ao λ_1 e λ_2 , respectivamente, então v_1 e v_2 são ortogonais, pois $\lambda_1 \cdot (v_1' \cdot v_2) = (\lambda_1 \cdot v_1)' \cdot v_2 = (v_1' \cdot A) \cdot v_2 = v_1' \cdot (A \cdot v_2) = v_1' \cdot (\lambda_2 \cdot v_2) = \lambda_2 \cdot (v_1' \cdot v_2)$, ou seja, $\lambda_1 \cdot (v_1' \cdot v_2) = \lambda_2 \cdot (v_1' \cdot v_2)$, como $\lambda_1 \neq \lambda_2$, então $v_1' \cdot v_2 = 0$. Podemos tomar v_1 e v_2 de comprimento 1 = $\|v_1\| = \|v_2\|$. Por exemplo, digamos que $v_1 = \begin{pmatrix} x \\ y \end{pmatrix}$, $x^2 + y^2 = 1$, então podemos tomar $v_2 = \begin{pmatrix} -y \\ x \end{pmatrix}$. Por que? A matriz $P = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ é uma matriz ortogonal e é matriz de troca de base de $\{v_1, v_2\}$ para a base canônica $\{e_1, e_2\}$. Sendo assim, no novo sistema de coordenadas representado por $\{v_1, v_2\}$:

$$\bar{A} = \begin{pmatrix} x & y \\ -y & x \end{pmatrix} \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} x & -y \\ y & x \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Nesse novo sistema, $\mathcal{Q}(\bar{x}, \bar{y}) = \lambda_1 \cdot (\bar{x})^2 + \lambda_2 \cdot (\bar{y})^2$.

Exemplo 2: Seja a forma quadrática $Q(x, y) = \frac{5}{4} \cdot x^2 + \frac{\sqrt{3}}{2} \cdot x \cdot y + \frac{7}{4} \cdot y^2 = (x \ y) \cdot \begin{pmatrix} \frac{5}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$. O polinômio característico da matriz $A = \begin{pmatrix} \frac{5}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} \end{pmatrix}$ é $p(\lambda) = \lambda^2 - 3 \cdot \lambda + 2 = (\lambda - 1) \cdot (\lambda - 2)$. Os autovalores da matriz A são $\lambda_1 = 1$ e $\lambda_2 = 2$. Portanto, a forma quadrática é positiva-definida. Se $v_1 = \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \cos\left(\frac{5 \cdot \pi}{6}\right) \\ \sin\left(\frac{5 \cdot \pi}{6}\right) \end{pmatrix}$ e $v_2 = \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix} = \begin{pmatrix} \cos\left(\frac{8 \cdot \pi}{6}\right) \\ \sin\left(\frac{8 \cdot \pi}{6}\right) \end{pmatrix}$, observe que $A \cdot v_1 = v_1$ e $A \cdot v_2 = 2 \cdot v_2$. Observe que $\begin{pmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{5}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} \end{pmatrix} \cdot \begin{pmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Se adotamos a base $B = \{v_1, v_2\}$ para o plano $\mathbb{R}^2$, então $P = \begin{pmatrix} -\frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}$ é a troca de base de $B = \{v_1, v_2\}$ para a base canônica $\{e_1, e_2\}$. Na base $B = \{v_1, v_2\}$, que corresponde ao sistema de coordenadas $\bar{x}, \bar{y}$, a expressão da forma é $Q(\bar{x}, \bar{y}) = 1 \cdot (\bar{x})^2 + 2 \cdot (\bar{y})^2$. Geometricamente, a base canônica rotou 150° no sentido anti-horário.



Observe que a definição da forma é determinada pelos sinais dos autovalores. A forma quadrática é positiva-definida se, e só se, $\lambda_1 > 0, \lambda_2 > 0$; negativa-definida se, e só se, $\lambda_1 < 0, \lambda_2 < 0$; positiva-semidefinida se, e só se, $\lambda_1 > 0, \lambda_2 = 0$; negativa-semidefinida se, e só se, $\lambda_1 = 0, \lambda_2 < 0$; indefinida se, e só se, $\lambda_1 > 0, \lambda_2 < 0$.

Como o polinômio característico de A é $p(\lambda) = \lambda^2 - (a + c) \cdot \lambda + a \cdot c - b^2$, facilmente verificamos que $\text{traço} A = \text{traço} \begin{pmatrix} a & b \\ b & c \end{pmatrix} = a + c = \lambda_1 + \lambda_2$ e $\det A = \begin{vmatrix} a & b \\ b & c \end{vmatrix} = a \cdot c - b^2 = \lambda_1 \cdot \lambda_2$, portanto, $p(\lambda) = \lambda^2 - (\text{traço} A) \cdot \lambda + |\det A|$. Logo, a forma quadrática é positiva-definida se, e só se, $\text{traço} A > 0, \det A > 0$; negativa-definida se, e só se, $\text{traço} A < 0, \det A > 0$; positiva-

semidefinida se, e só se, $\text{traço} A > 0, \det A = 0$; negativa-semidefinida se, e só se, $\text{traço} A < 0, \det A = 0$; indefinida se, e só se, $\det A < 0$.

Exemplo 3: A expressão de uma forma como uma soma de quadrados não é única. Por exemplo, $Q(x, y) = \frac{5}{4} \cdot x^2 + \frac{\sqrt{3}}{2} \cdot x \cdot y + \frac{7}{4} \cdot y^2 = \frac{5}{4} \cdot \left(x + \frac{\sqrt{3}}{5} \cdot y\right)^2 + \frac{8}{5} \cdot y^2$. Seja a troca de coordenadas $\bar{x} = x + \frac{\sqrt{3}}{5} \cdot y, \bar{y} = y$ dada pela matriz $L = \begin{pmatrix} 1 & \frac{\sqrt{3}}{5} \\ 0 & 1 \end{pmatrix}$. Observe

que $L^{-1} = \begin{pmatrix} 1 & -\frac{\sqrt{3}}{5} \\ 0 & 1 \end{pmatrix}$, então $(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} 1 & 0 \\ -\frac{\sqrt{3}}{5} & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{5}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} \end{pmatrix} \cdot \begin{pmatrix} 1 & -\frac{\sqrt{3}}{5} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{5}{4} & 0 \\ 0 & \frac{8}{5} \end{pmatrix} = \bar{A}$.

Através da troca de coordenadas $\bar{x} = \frac{\sqrt{5}}{2} \cdot x + \frac{\sqrt{15}}{10} \cdot y, \bar{y} = \frac{2 \cdot \sqrt{10}}{5} \cdot y$ dada pela matriz $L = \begin{pmatrix} \frac{\sqrt{5}}{2} & \frac{\sqrt{15}}{10} \\ 0 & \frac{2 \cdot \sqrt{10}}{5} \end{pmatrix}$. Observe que $L^{-1} = \begin{pmatrix} \frac{2 \cdot \sqrt{5}}{5} & -\frac{\sqrt{30}}{20} \\ 0 & \frac{\sqrt{10}}{4} \end{pmatrix}$, então $(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} \frac{2 \cdot \sqrt{5}}{5} & 0 \\ -\frac{\sqrt{30}}{20} & \frac{\sqrt{10}}{4} \end{pmatrix} \cdot \begin{pmatrix} \frac{5}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} \end{pmatrix} \cdot \begin{pmatrix} \frac{2 \cdot \sqrt{5}}{5} & -\frac{\sqrt{30}}{20} \\ 0 & \frac{\sqrt{10}}{4} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \bar{A}$. Observe que $\begin{pmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & \frac{\sqrt{5}}{\sqrt{8}} \end{pmatrix} \cdot \begin{pmatrix} \frac{5}{4} & 0 \\ 0 & \frac{8}{5} \end{pmatrix} \cdot \begin{pmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & \frac{\sqrt{5}}{\sqrt{8}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ e descubra como obtivemos a troca de coordenadas acima.

Exercício 1

Seja a forma quadrática $Q(x, y) = 4 \cdot x^2 + 4 \cdot x \cdot y + 2 \cdot y^2 = (2 \cdot x + y)^2 + y^2 = (x \ y) \cdot \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$. Qual é a matriz da forma acima no sistema de coordenadas $\bar{x} = 2 \cdot x + y, \bar{y} = y$? Observe que o fato de a forma ser representada por uma matriz diagonal $\begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}$ não implica que os autovalores da matriz da forma na base canônica, no caso $\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$, sejam os elementos na diagonal da matriz $\begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}$. $Q(\bar{x}, \bar{y}) = (\bar{x} \ \bar{y}) \cdot \begin{pmatrix} 1/2 & 0 \\ -1/2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 & -1/2 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \bar{x}^2 + \bar{y}^2$. Os autovalores da matriz $\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$ são $3 \pm \sqrt{5}$.

Exemplo 4: Seja a forma quadrática $Q(x, y) = x^2 + 3 \cdot x \cdot y + y^2 = (x \ y) \cdot \begin{pmatrix} 1 & 3/2 \\ 3/2 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \left(x + \frac{3}{2} \cdot y\right)^2 - \left(\frac{\sqrt{5}}{2} \cdot y\right)^2$. Os autovalores de $\begin{pmatrix} 1 & 3/2 \\ 3/2 & 1 \end{pmatrix}$ são $\lambda_1 = -\frac{1}{2}$ e $\lambda_2 = \frac{5}{2}$.

$(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} 1 & 0 \\ -\frac{3 \cdot \sqrt{5}}{5} & \frac{2 \cdot \sqrt{5}}{5} \end{pmatrix} \cdot \begin{pmatrix} \frac{5}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{7}{4} \end{pmatrix} \cdot \begin{pmatrix} 1 & -\frac{3 \cdot \sqrt{5}}{5} \\ 0 & \frac{2 \cdot \sqrt{5}}{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \bar{A}$, $\begin{pmatrix} 1 & -\frac{3 \cdot \sqrt{5}}{5} \\ 0 & \frac{2 \cdot \sqrt{5}}{5} \end{pmatrix}$ é a inversa da matriz $\begin{pmatrix} 1 & \frac{3}{2} \\ 0 & \frac{\sqrt{5}}{2} \end{pmatrix}$.

Exemplo 5: Seja a forma quadrática $Q(x,y) = x^2 - 2 \cdot x \cdot y + y^2 = (x \ y) \cdot \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = (x-y)^2 + 0 \cdot y^2$. Os autovalores de $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ são $\lambda_1 = 0$ e $\lambda_2 = 2$. $(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \bar{A}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ é a inversa da matriz $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.

Exemplo 6: Seja a forma quadrática $Q(x,y) = x^2 + 2 \cdot x \cdot y + y^2 = (x \ y) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = (x+y)^2 + 0 \cdot y^2$. Os autovalores de $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ são $\lambda_1 = 0$ e $\lambda_2 = 2$. O fato que os autovalores de duas matrizes sejam os mesmos não implica que estejamos com a mesma forma quadrática: $x^2 - 2 \cdot x \cdot y + y^2$ não é $x^2 + 2 \cdot x \cdot y + y^2$. $(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \bar{A}$, $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ é a inversa da matriz $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Exemplo 7: Se $Q(x,y) = -x^2 - 2 \cdot x \cdot y - y^2 = (x \ y) \cdot \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = -(x+y)^2 + 0 \cdot y^2$. Os autovalores de $\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$ são $\lambda_1 = 0$ e $\lambda_2 = -2$. $(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} = \bar{A}$.

Exemplo 8: Seja a forma quadrática $Q(x,y) = -5 \cdot x^2 + 8 \cdot x \cdot y - 5 \cdot y^2 = (x \ y) \cdot \begin{pmatrix} -5 & 4 \\ 4 & -5 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = -\left(\sqrt{5} \cdot x - \frac{4 \cdot \sqrt{5}}{5} \cdot y\right)^2 - \left(\frac{3 \cdot \sqrt{5}}{5} \cdot y\right)^2$. Os autovalores de $\begin{pmatrix} -5 & 4 \\ 4 & -5 \end{pmatrix}$ são $\lambda_1 = -9$ e $\lambda_2 = -1$. $(L^{-1})' \cdot A \cdot (L^{-1}) = \begin{pmatrix} \frac{\sqrt{5}}{5} & 0 \\ \frac{4 \cdot \sqrt{5}}{15} & \frac{\sqrt{5}}{3} \end{pmatrix} \cdot \begin{pmatrix} -5 & 4 \\ 4 & -5 \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{5}}{5} & \frac{4 \cdot \sqrt{5}}{15} \\ 0 & \frac{\sqrt{5}}{3} \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \bar{A}$, $\begin{pmatrix} \frac{\sqrt{5}}{5} & \frac{4 \cdot \sqrt{5}}{15} \\ 0 & \frac{\sqrt{5}}{3} \end{pmatrix}$ é a inversa da matriz $\begin{pmatrix} \sqrt{5} & -\frac{4 \cdot \sqrt{5}}{5} \\ 0 & \frac{3 \cdot \sqrt{5}}{5} \end{pmatrix}$.

Portanto, na redução de uma forma quadrática a uma soma de quadrados, não interessam os valores dos autovalores, mas sim, seus sinais: se os dois autovalores são positivos, então pode ser reduzida à matriz $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, se um autovalor é negativo e outro positivo, pode ser reduzida à matriz $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, se ambos os autovalores são negativos, pode ser reduzida à matriz $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$; finalmente, se um autovalor é positivo e o outro nulo ou se um autovalor é negativo e o outro nulo, a forma pode ser reduzida às matrizes $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ e $\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$, respectivamente.

Classificamos todas as formas quadráticas em $\mathbb{R}^2$. Elas se reduzem às positivas-definidas, para as quais existe uma base do $\mathbb{R}^2$, na qual a matriz da forma é $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$; às negativas-definidas, para as quais existe uma base do $\mathbb{R}^2$, na qual a matriz da forma é $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$; às positivas-semidefinidas, para as quais existe uma base do $\mathbb{R}^2$, na qual a matriz

da forma é $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, às negativas-semidefinidas, para as quais existe uma base do $\mathbb{R}^2$, na qual a matriz da forma é $\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$; finalmente, às indefinidas, para as quais existe uma base do $\mathbb{R}^2$, na qual a matriz da forma é $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Estendendo o resultado acima para formas no $\mathbb{R}^n$, uma argumentação similar implica que sempre existe uma base no $\mathbb{R}^n$ na qual a matriz da forma quadrática $x' \cdot A \cdot x$ se reduz a

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \hline & & & -1 \\ & & & \ddots \\ & & & & -1 \\ \hline & & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix}$$

$\underbrace{\hspace{10em}}_{\pi + \nu + (n - \pi - \nu)}$

tal que o número de 1's corresponde ao número π de autovalores positivos da matriz A , o número ν de -1's corresponde ao número de autovalores negativos e os zeros correspondem aos autovalores nulos e são $n - \pi - \nu$ "zeros". A diferença σ entre o número π de autovalores positivos e o número ν de autovalores negativos é chamada de "assinatura da forma". Observe que $\pi + \nu$ é o posto da matriz A e $n - (\pi + \nu)$ é o número de autovalores nulos.

Em 1852, o matemático inglês James Joseph Sylvester (1814-1897) publicou o resultado conhecido como "Lei de Inércia para Formas Quadráticas"³. Esse resultado assegura que qualquer que seja a representação da forma quadrática em soma de quadrados, $Q = c_1 \cdot x_1^2 + c_2 \cdot x_2^2 + \dots + c_n \cdot x_n^2$, os números de coeficientes c_i positivos, negativos e nulos são invariantes, ou seja, π , ν e $n - \pi - \nu$, respectivamente. O resultado principal do artigo é a redução de uma forma quadrática a uma soma de quadrados através de uma troca de coordenadas ortogonais, que é o teorema da diagonalização de matrizes simétricas apresentado nos cursos de Álgebra Linear.

Se a forma é positiva-definida se, e só se, $\pi = n$, a forma é negativa-definida se, e só se, $\nu = n$; a forma é positiva-semidefinida se, e só se, $\pi < n$ e $\nu = 0$, a forma é negativa-semidefinida se, e só se, $\pi = 0$ e $\nu < n$ e a forma é indefinida se, e só se, $0 < \pi$ e $0 < \nu$. Ainda, a forma é positiva-definida se, e só se, $\sigma = n$, a forma é negativa-definida se, e só se, $\sigma = -n$; a forma é positiva-semidefinida se, e só se, $\sigma = \pi < n$, a forma é negativa-semidefinida se, e só se, $-\sigma = \nu < n$ e a forma é indefinida se, e só se, $0 < \pi$ e $0 < \nu$.

Exercício 2

Usando que $\text{traço} A$ e $\det A$, determine a definição das formas abaixo:

i) $q(x, y) = 4 \cdot x^2 - 24 \cdot x \cdot y + 25 \cdot y^2$;

ii) $q(x, y) = 6 \cdot x^2 - 24 \cdot x \cdot y + 24 \cdot y^2$;

iii) $q(x, y) = 12 \cdot x^2 + 12 \cdot x \cdot y + 7 \cdot y^2$;

iv) $q(x, y) = 4 \cdot x^2 + 2 \cdot x \cdot y + y^2$;

v) $q(x, y) = 5 \cdot x^2 + 2 \cdot x \cdot y + 10 \cdot y^2$;

vi) $q(x, y) = x^2 - 4 \cdot x \cdot y + y^2$.

4. A derivada segunda $\nabla^2 f$

Nossas conclusões em lições anteriores resultaram das informações obtidas através da análise dos planos tangentes aos gráficos das funções. O lagrangeano é apenas uma maneira de dizer que, num ponto de mínimo ou máximo da função, necessariamente, a isoquanta à qual esse ponto pertence tangencia a superfície determinada pelas restrições.

Uma função da forma $q(x, y) = a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2 + d \cdot x + e \cdot y + f$ é dita uma quádrlica e matricialmente, é representada por $q(x, y) = \begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d & e \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + f = 0$. Uma cônica é uma curva no plano determinada por um polinômio de

grau 2, isto é, $a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2 + d \cdot x + e \cdot y + f = 0$. Observe que a cônica é a isoquanta de nível 0 da função q .

Para as funções $f(x, y)$ definidas em $\mathbb{R}^2$, o plano tangente $z = \frac{\partial f}{\partial x}(x^*, y^*) \cdot (x - x^*) + \frac{\partial f}{\partial y}(x^*, y^*) \cdot (y - y^*) + f(x^*, y^*) = d \cdot x + e \cdot y + f$, associado à derivada primeira $\nabla f(x^*, y^*)$, corresponde à parte linear $d \cdot x + e \cdot y + f$ da quádrlica. A soma dos termos de grau dois, $a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2$, corresponde à derivada $\nabla^2 f(x^*, y^*)$, como veremos a seguir.

Se, anteriormente, aproximamos as funções pelos planos tangentes $z = \frac{\partial f}{\partial x}(x^*, y^*) \cdot (x - x^*) + \frac{\partial f}{\partial y}(x^*, y^*) \cdot (y - y^*) + f(x^*, y^*) = d \cdot x + e \cdot y + f$, a partir de agora, as aproximaremos pelo plano tangente somado à parte quadrática $(x - x^*, y - y^*)' \cdot \nabla^2 f(\bar{x}, \bar{y}) \cdot (x - x^*, y - y^*) = a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2$, em que $(\bar{x}, \bar{y})$ é um ponto intermediário sobre o segmento de reta que une (x, y) e (x^*, y^*) . Portanto, ajustaremos às nossas funções, como aproximações, um parabolóide elíptico ou hiperbólico ou um cilindro parabólico. Como essas são todas as possíveis superfícies da forma $z = a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2 + d \cdot x + e \cdot y + f$ no espaço $\mathbb{R}^3$, obtemos em princípio, através desse processo, uma classificação das funções nas vizinhanças de um ponto qualquer (x^*, y^*) . Dizemos “em princípio” porque muitas funções não se ajustam a essa classificação. Na vizinhança de um ponto, se possível, para obter a função, precisamos “deformar” o plano tangente de maneira a resultar num parabolóide ou num cilindro. Essas são as duas razões pelas quais iniciamos com as quádrlicas $z = a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2 + d \cdot x + e \cdot y + f$ e as respectivas isoquantas $a \cdot x^2 + 2 \cdot b \cdot x \cdot y + c \cdot y^2 + d \cdot x + e \cdot y + f = g$.

Primeiramente, vejamos na reta. Seja uma função de variável real $f(x)$ com derivadas $f'(x), f''(x)$ contínuas. Definamos a função $h(t) = (x - x_0)^2 \cdot (f(x) - f(t) - f'(t) \cdot (x - t)) - (f(x) - f(x_0) - f'(x_0) \cdot (x - x_0)) \cdot (x - t)^2$. Observe que $h(x_0) = h(x) = 0$. Portanto, necessariamente, $h(t)$ tem um ponto de derivada nula entre x_0 e x , pois se $h(t)$ é nula em todo o intervalo entre x_0 e x , então todo o ponto t entre x_0 e x é tal que $h'(t) = 0$; se $h(t)$ não é nula, tem um ponto $\bar{t}$ de mínimo ou de máximo no intervalo entre x_0 e x e, nesse caso, $h'(\bar{t}) = 0$.

Como $h'(\bar{t}) = (x - x_0)^2 \cdot (-f''(\bar{t}) \cdot (x - \bar{t})) + 2 \cdot (f(x) - f(x_0) - f'(x_0) \cdot (x - x_0)) \cdot (x - \bar{t}) = 0$, então $f(x) - f(x_0) - f'(x_0) \cdot (x - x_0) = \frac{f''(\bar{t}) \cdot (x - x_0)^2}{2}$. Sendo assim, concluímos que $f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(\bar{t})}{2} \cdot (x - x_0)^2$, para algum $\bar{t}$ entre x_0 e x , isto é, $x_0 < \bar{t} < x$ ou $x < \bar{t} < x_0$. $R(x_0, \bar{t}(x_0, x), x) = \frac{f''(\bar{t})}{2} \cdot (x - x_0)^2$ é chamado **Resto na Forma de Lagrange**, $\bar{t} = \bar{t}(x_0, x)$ depende de x_0 e x . Observe que $\lim_{x \rightarrow x_0} \frac{R(x_0, \bar{t}, x)}{(x - x_0)} = \lim_{x \rightarrow x_0} \frac{f''(\bar{t})}{2} \cdot (x - x_0) = 0$.

O resultado acima permite-nos estabelecer uma conclusão muito importante. Digamos que $f'' \geq 0$ em determinado intervalo, desse modo, em qualquer $\bar{t}$ nesse intervalo, $f''(\bar{t}) \geq 0$, portanto, $f(x) \geq f(x_0) + f'(x_0) \cdot (x - x_0)$ e concluímos que f é convexa. Por outro lado, se f é convexa no intervalo, que pode ser apenas à direita ou à esquerda de x^* , e existir um ponto x^* tal que $f''(x^*) < 0$, então, pela continuidade de f'' , existe um subintervalo que pode ser apenas à direita ou à esquerda de x^* , no qual $f'' < 0$, portanto, existem $x_0, \bar{t}, x$ na proximidade de x^* , tais que $f(x) - (f(x_0) + f'(x_0) \cdot (x - x_0)) = \frac{f''(\bar{t}) \cdot (x - x_0)}{2} < 0$. Nesse caso, $f(x) < f(x_0) + f'(x_0) \cdot (x - x_0)$ e f não é convexa. Em resumo, f é convexa num intervalo se, e só se, $f'' \geq 0$ no mesmo intervalo. Da mesma maneira, apenas invertendo a direção das desigualdades, mostramos que f é côncava num intervalo se, e só se, $f'' \leq 0$ no mesmo intervalo.

Caso $f'' > 0$ num certo intervalo, então f é estritamente convexa nesse intervalo, e se $f'' < 0$ num certo intervalo, então f é estritamente côncava nesse intervalo. Se f é estritamente convexa num intervalo, então não é verdade que $f'' > 0$ nesse intervalo. Observe o exemplo de $f(x) = x^4$ na reta. Essa função é estritamente convexa, mas não é verdade que $f'' > 0$ na reta, pois $f''(0) = 0$. Da mesma maneira, se f é estritamente côncava num intervalo, não implica $f'' < 0$ nesse intervalo. Observe $f(x) = -x^4$.

Os resultados acima são tão importantes que os destacamos. Seja I um intervalo qualquer da reta.

Teorema 1

i) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ é **convexa** se, e somente se, $f'' \geq 0$ em I ;

- ii) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ é **côncava** se, e somente se, $f'' \leq 0$ em I ;
- iii) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ se $f'' > 0$ em I , então f é **estritamente convexa** em I ;
- iv) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ se $f'' < 0$ em I , então f é **estritamente côncava** em I .

Não é difícil perceber que

Teorema 2

- i) $f'(x^*) = 0$ e $f''(x^*) > 0$, então x^* é um mínimo local;
- ii) $f'(x^*) = 0$ e $f''(x^*) < 0$, então x^* é um máximo local.

Para estabelecer esse último resultado, observamos que se $f''(x^*) > 0$, então localmente, ou seja, na proximidade de x^* , temos $f'' > 0$, ou seja, f é localmente convexa. Da mesma maneira, se $f''(x^*) < 0$, então localmente, ou seja, na proximidade de x^* , temos $f'' < 0$, ou seja, f é localmente côncava.

O esforço seguinte é obter a derivada segunda para uma função de várias variáveis e estabelecer os dois teoremas acima para essas funções.

Não precisamos que a função $f: S \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ esteja definida num conjunto convexo desde que em cada ponto de $x \in S$ tenhamos uma bola aberta $B(x, \varepsilon) = \{y \in \mathbb{R}^2 \mid \|x - y\| < \varepsilon\} \subseteq S$, mas, por facilidade, admitamos que S seja convexo. Seja uma função $f: S \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, $S \subseteq \mathbb{R}^2$ subconjunto convexo sejam $(x_1^*, x_2^*), (x_1, x_2) \in S$. O segmento de reta unindo (x_1^*, x_2^*) e (x_1, x_2) é $(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*))$, para $0 \leq t \leq 1$, e está contido em S . Seja a função $\varphi(t) = f(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*))$. Pela regra da cadeia,

$$\varphi'(t) = \nabla f(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*)) \cdot \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{pmatrix}, \quad \text{isto é,}$$

$$\varphi'(t) = \frac{\partial f}{\partial x_1}(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*)) \cdot (x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*)) \cdot (x_2 - x_2^*).$$

Aplicando a regra da cadeia novamente, obtemos

$$\varphi''(t) = \left(\frac{\partial^2 f}{\partial x_1^2} \cdot (x_1 - x_1^*) + \frac{\partial^2 f}{\partial x_2 \partial x_1} \cdot (x_2 - x_2^*) \right) \cdot (x_1 - x_1^*) + \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \cdot (x_1 - x_1^*) + \frac{\partial^2 f}{\partial x_2^2} \cdot (x_2 - x_2^*) \right) \cdot (x_2 - x_2^*).$$

Retiramos o argumento $(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*))$ das derivadas parciais por óbvia comodidade. Pelo que acabamos de ver: $\varphi(t) = \varphi(0) + \varphi'(0) \cdot (t - 0) + \frac{\varphi''(\bar{t})}{2} \cdot (t - 0)^2$, para algum $\bar{t}$ entre 0 e t . Fazendo $t = 1$, obtém-se $\varphi(1) = \varphi(0) + \varphi'(0) + \frac{\varphi''(\bar{t})}{2}$, para

algum $\bar{t}$ entre 0 e 1. Portanto, lembrando que $\varphi(t) = f(x_1^* + t \cdot (x_1 - x_1^*), x_2^* + t \cdot (x_2 - x_2^*))$, concluímos que:

$$\begin{aligned} f(x_1, x_2) = & f(x_1^*, x_2^*) + \frac{\partial f}{\partial x_1}(x_1^*, x_2^*) \cdot (x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \cdot (x_2 - x_2^*) + \\ & \frac{1}{2} \cdot \left(\frac{\partial^2 f}{\partial x_1^2}(\bar{x}_1, \bar{x}_2) \cdot (x_1 - x_1^*)^2 + \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}_1, \bar{x}_2) \cdot (x_2 - x_2^*) \cdot (x_1 - x_1^*) + \right. \\ & \left. \frac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{x}_1, \bar{x}_2) \cdot (x_1 - x_1^*) \cdot (x_2 - x_2^*) + \frac{\partial^2 f}{\partial x_2^2}(\bar{x}_1, \bar{x}_2) \cdot (x_2 - x_2^*)^2 \right). \end{aligned}$$

Perceba que $\varphi(0) + \varphi'(0) = f(x_1^*, x_2^*) + \frac{\partial f}{\partial x_1}(x_1^*, x_2^*) \cdot (x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \cdot (x_2 - x_2^*)$ corresponde ao plano tangente do gráfico de f no ponto $(x_1^*, x_2^*, f(x_1^*, x_2^*))$, e que:

$$\begin{aligned} \frac{\varphi''(\bar{t})}{2} = & \frac{1}{2} \cdot \left(\frac{\partial^2 f}{\partial x_1^2}(\bar{x}_1, \bar{x}_2) \cdot (x_1 - x_1^*)^2 + \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}_1, \bar{x}_2) \cdot (x_2 - x_2^*) \cdot (x_1 - x_1^*) + \right. \\ & \left. \frac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{x}_1, \bar{x}_2) \cdot (x_1 - x_1^*) \cdot (x_2 - x_2^*) + \frac{\partial^2 f}{\partial x_2^2}(\bar{x}_1, \bar{x}_2) \cdot (x_2 - x_2^*)^2 \right). \end{aligned}$$

Observe que $\frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}_1, \bar{x}_2) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{x}_1, \bar{x}_2)$. Se $\bar{t}$ está entre 0 e 1, então $(x_1^* + \bar{t} \cdot (x_1 - x_1^*), x_2^* + \bar{t} \cdot (x_2 - x_2^*)) = (\bar{x}_1, \bar{x}_2)$ é um ponto entre (x_1^*, x_2^*) e (x_1, x_2) . A expressão acima representada matricialmente é:

$$\frac{1}{2} \cdot \begin{pmatrix} x_1 - x_1^* & x_2 - x_2^* \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\bar{x}_1, \bar{x}_2) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}_1, \bar{x}_2) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{x}_1, \bar{x}_2) & \frac{\partial^2 f}{\partial x_2^2}(\bar{x}_1, \bar{x}_2) \end{pmatrix} \cdot \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{pmatrix},$$

que é uma forma quadrática nas variáveis $x_1 - x_1^*$ e $x_2 - x_2^*$.

Dada a função $f: S \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, definimos a derivada segunda de f no ponto (x_1, x_2) como a matriz:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_2^2}(x_1, x_2) \end{pmatrix}.$$

Usando uma notação mais compacta, obtemos:

$$f(x_1, x_2) = f(x_1^*, x_2^*) + \nabla f(x_1^*, x_2^*) \cdot \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{pmatrix} + \frac{1}{2} \cdot \begin{pmatrix} x_1 - x_1^* & x_2 - x_2^* \end{pmatrix} \cdot \nabla^2 f(\bar{x}) \cdot \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{pmatrix}$$

Se $(x_1^*, x_2^*) = x^*$ e $(x_1, x_2) = x$, então $f(x) = f(x^*) + \nabla f(x^*) \cdot (x - x^*) + \frac{1}{2} \cdot (x - x^*)' \cdot \nabla^2 f(\bar{x}) \cdot (x - x^*)$, em que, como antes, $\nabla f(x^*) \cdot (x - x^*)$ é o produto escalar do gradiente $\nabla f(x^*)$ com o vetor $(x - x^*)$,

e $(x-x^*)'$ é o transposto do vetor $(x-x^*) = \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{pmatrix}$.

Dada uma matriz $A, n \times n$, e um vetor $x \in \mathbb{R}^n$, $A \cdot x = \begin{pmatrix} a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n \\ \vdots \\ a_{n1} \cdot x_1 + \dots + a_{nn} \cdot x_n \end{pmatrix} = \begin{pmatrix} A'_1 \cdot x \\ \vdots \\ A'_n \cdot x \end{pmatrix}$, em que $A'_i = (a_{i1} \dots a_{in})$, $i = 1, \dots, n$ são as linhas da matriz A , então $\|A \cdot x\| \leq |A'_1 \cdot x| + \dots + |A'_n \cdot x| \leq (\|A_1\| + \dots + \|A_n\|) \cdot \|x\| = K \cdot \|x\|$ (ver lição 1). Portanto, dada uma forma quadrática $x' \cdot A \cdot x$, $|x' \cdot A \cdot x| \leq \|x'\| \cdot \|A \cdot x\| \leq K \cdot \|x\|^2$ para K , tal que $K \geq \|A_1\| + \dots + \|A_n\|$. A partir disso, como anteriormente, $R = R(x^*, \bar{x}(x^*, x), x) = \frac{1}{2} \cdot (x-x^*)' \cdot \nabla^2 f(\bar{x}) \cdot (x-x^*)$, o resto na forma de Lagrange, é tal que $\lim_{x \rightarrow x^*} \frac{R}{\|x-x^*\|} = \lim_{x \rightarrow x^*} \frac{1}{2} \cdot \frac{|(x-x^*)' \cdot \nabla^2 f(\bar{x}) \cdot (x-x^*)|}{\|x-x^*\|} = 0$.

Não é difícil perceber que o mesmo desenvolvimento pode se estender para uma função $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, em que $S \subseteq \mathbb{R}^n$ é convexo. Dados $(x_1^*, \dots, x_n^*) = x^* \in S$ e $(x_1, \dots, x_n) = x \in S$, tomemos o segmento de reta unindo x^* e x , $x^* + t \cdot (x-x^*)$, para $0 \leq t \leq 1$, e definamos $\phi(t) = f(x^* + t \cdot (x-x^*))$. E, assim, $\phi'(t) = \nabla f(x^* + t \cdot (x-x^*)) \cdot (x-x^*)$. De maneira idêntica, aplicando a regra da cadeia em $\phi'(t) = \nabla f(x^* + t \cdot (x-x^*)) \cdot (x-x^*)$, obtemos, como anteriormente,

$$\begin{aligned} \phi''(t) &= \frac{\partial^2 f}{\partial x_1^2}(x^* + t \cdot (x-x^*)) \cdot (x_1 - x_1^*)^2 + \dots + \frac{\partial^2 f}{\partial x_n^2}(x^* + t \cdot (x-x^*)) \cdot (x_n - x_n^*)^2 + \\ &+ 2 \cdot \sum_{i < j} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^* + t \cdot (x-x^*)) \cdot (x_i - x_i^*) \cdot (x_j - x_j^*), \end{aligned}$$

usando que $\frac{\partial^2 f}{\partial x_i \partial x_j}(x^* + t \cdot (x-x^*)) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x^* + t \cdot (x-x^*))$. Da mesma maneira que anteriormente, a expressão acima expressa matricialmente resulta

$$(x_1 - x_1^* \quad \dots \quad x_n - x_n^*) \cdot \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^* + t \cdot (x-x^*)) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^* + t \cdot (x-x^*)) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^* + t \cdot (x-x^*)) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x^* + t \cdot (x-x^*)) \end{pmatrix} \cdot \begin{pmatrix} x_1 - x_1^* \\ \vdots \\ x_n - x_n^* \end{pmatrix}.$$

Dada uma função $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, definimos a derivada segunda de f em x como a matriz:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_3 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial^2 x_2}(x) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_3}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(x) & \frac{\partial^2 f}{\partial^2 x_3}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_3}(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_3 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial^2 x_n}(x) \end{pmatrix}.$$

A matriz $\nabla^2 f(x)$ é dita matriz hessiana de f em x . Observe que $\nabla^2 f(x)$ é uma matriz simétrica, pois $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$, e que, portanto, define a forma quadrática

$$v' \cdot \nabla^2 f(x) \cdot v = (v_1 \quad \cdots \quad v_n) \cdot \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_3 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial^2 x_2}(x) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_3}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(x) & \frac{\partial^2 f}{\partial^2 x_3}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_3}(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_3 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial^2 x_n}(x) \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Dada uma função $f: S \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, $S \subseteq \mathbb{R}^2$ subconjunto convexo, com derivadas $\nabla f(x), \nabla^2 f(x)$, então $f(x) = f(x^*) + \nabla f(x^*) \cdot (x - x^*) + \frac{1}{2} \cdot (x - x^*)' \cdot \nabla^2 f(\bar{x}) \cdot (x - x^*)$, para algum $\bar{x}$ no segmento de reta $\lambda \cdot x + (1 - \lambda) \cdot x^*, \lambda \in (0, 1)$, isto é, $\bar{x} = \bar{\lambda} \cdot x + (1 - \bar{\lambda}) \cdot x^*$, para algum $\bar{\lambda}$, tal que $0 < \bar{\lambda} < 1$. $R = \frac{1}{2} \cdot (x - x^*)' \cdot \nabla^2 f(\bar{x}) \cdot (x - x^*)$ é o resto na forma de Lagrange e que

$$\lim_{x \rightarrow x^*} \frac{R}{\|x - x^*\|} = \lim_{x \rightarrow x^*} \frac{1}{2} \cdot \frac{\left| (x - x^*)' \cdot \nabla^2 f(\bar{x}) \cdot (x - x^*) \right|}{\|x - x^*\|} = 0.$$

Suponhamos que $S \subseteq \mathbb{R}^n$ é um subconjunto convexo. As demonstrações dos teoremas abaixo são as mesmas feitas anteriormente.

Teorema 3

- i) $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa se, e somente se, $\nabla f = f''$ é positiva semidefinida em S ;
- ii) $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é côncava se, e somente se, $\nabla^2 f = f''$ é negativa semidefinida em S ;

- iii) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, se $\nabla^2 f = f''$ é positiva definida em s , então f é estritamente convexa em s ;
- iv) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, se $\nabla^2 f = f''$ é negativa definida em s , então f é estritamente côncava em s .

Teorema 4

- i) $\nabla f(x^*) = f'(x^*) = 0$ e $\nabla^2 f(x^*) = f''(x^*)$ é positiva-definida, então x^* é um mínimo local;
- ii) $\nabla f(x^*) = f'(x^*) = 0$ e $\nabla^2 f(x^*) = f''(x^*)$ é negativa-definida, então x^* é um máximo local.

Exemplo 9: Uma primeira questão a esclarecer é se no teorema acima não podemos substituir positiva-definida e negativa-definida por positiva-semidefinida e negativa-semidefinida, respectivamente. A resposta é não! Consideremos a função $f(x, y) = x^2 - y^4$. Obviamente, o ponto $(0, 0)$ não é nem mínimo e nem máximo. $f''(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & -12 \cdot y^2 \end{pmatrix}$ são formas indefinidas se $y \neq 0$; $f''(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ é positiva-semidefinida. Portanto, não é verdade que $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ positiva-semidefinida impliquem que x^* é um mínimo local. A função $f(x, y) = -x^2 + y^4$ ilustra a possibilidade $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ negativa-semidefinida e $(0, 0)$ não sendo nem mínimo e nem máximo.

No resultado anterior, o que supomos é que f é localmente convexa, ou seja, convexa na proximidade de x^* , então, x^* é um mínimo local. Na parte ii), supomos que f é localmente côncava.

Exemplo 10: Consideremos a função $f(x, y) = x^4 + 2 \cdot x^2 \cdot y^2 + y^4 = (x^2 + y^2)^2$. As $f''(x, y) = \begin{pmatrix} 12 \cdot x^2 + 4 \cdot y^2 & 8 \cdot x \cdot y \\ 8 \cdot x \cdot y & 4 \cdot x^2 + 12 \cdot y^2 \end{pmatrix}$ são positivas-definidas, $(x, y) \neq (0, 0)$. $f''(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ não é positiva-semidefinida e sim nula! O ponto $(0, 0)$ é um mínimo para f . Esse resultado é óbvio e não precisamos da derivada segunda para obtê-lo. Mas, o exemplo ilustra que “ x^* é um mínimo local” não implica “ $f''(x^*)$ é positiva-semidefinida”. A razão deste fato é que ao

estabelecermos, através da definição das formas, que um ponto é mínimo local, não estamos empregando $f''(x^*)$ e sim $f''(\bar{x})$ para algum $\bar{x} \neq x^*$ na proximidade de x^* : $f(x,y) - f(0,0) = \frac{1}{2} \cdot (x \ y) \cdot \begin{pmatrix} 12 \cdot \bar{x}^2 + 4 \cdot \bar{y}^2 & 8 \cdot \bar{x} \cdot \bar{y} \\ 8 \cdot \bar{x} \cdot \bar{y} & 4 \cdot \bar{x}^2 + 12 \cdot \bar{y}^2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} > 0$, se $(x,y) \neq (0,0)$ e $(\bar{x},\bar{y})$ entre $(0,0)$ e (x,y) , portanto, $(\bar{x},\bar{y}) \neq (0,0)$.

5. A determinação da definição da forma quadrática

Através dos autovalores podemos determinar a definição de $\nabla^2 f(x) = A = (a_{ij})$, mas essa é uma tarefa difícil. É fácil para o caso $n=2$, pois o polinômio característico é $p(\lambda) = \lambda^2 - (a_{11} + a_{22}) \cdot \lambda + (a_{11} \cdot a_{22} - a_{12} \cdot a_{21})$, mas para $n \geq 3$ dependemos de particularidades da matriz A . O polinômio característico $p(\lambda)$ é de grau $n \geq 3$ e obter as raízes de tais polinômios, no caso geral, não é tarefa fácil. Por essa razão, vamos procurar outras maneiras de determinar a definição da forma quadrática sem ter que achar as raízes. De fato, não precisamos do valor das raízes, apenas precisamos saber se são negativas, nulas ou positivas. Buscar as raízes para saber sobre o sinal das mesmas parece um procedimento pouco econômico. Nas próximas seções responderemos a essa questão.

Existem muitas maneiras de reduzir uma forma quadrática a uma soma de quadrados. Todas essas reduções permitem determinar a definição da forma. Por exemplo, se todos os coeficientes c_i são positivos, a forma é positiva-definida; se algum c_i é zero, a forma é semidefinida; se alguns são positivos, outros, negativos, a forma é indefinida.

Todavia, como estamos exclusivamente interessados em saber a definição da forma, é muito mais razoável investigarmos se existem trocas de coordenadas que resultem nos coeficientes c_i sem que tenhamos que, efetivamente, fazer a redução. Esse procedimento é conhecido como método de Jacobi. Esse método, que apresentaremos mais adiante, permite determinar a definição da

forma através dos chamados determinantes menores principais da matriz A dos coeficientes.

5.1 Menores principais

Dada uma matriz $A = (a_{ij})$, $n \times n$, o **subdeterminante menor principal líder** A_k de ordem $k = 1, 2, \dots, n$ de A é obtido eliminando as $(n-k)$ últimas linhas e respectivas colunas de A . Se $k = n$, nenhuma linha e coluna é eliminada e $|A_n|$ é o determinante da matriz A . Portanto, as matrizes principais líderes são:

$$|A_1| = a_{11}, |A_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, |A_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \dots, |A_n| = |A|.$$

Designamos os $|A_k|$ subdeterminantes menores principais líderes, por menores principais líderes.

Um **subdeterminante menor principal** de ordem $k = 1, 2, \dots, n$ é obtido eliminando $(n-k)$ linhas e respectivas colunas de A . Se conservarmos as linhas e as colunas $i_1 < i_2 < \dots < i_k$ de A , então

$$|A(i_1, i_2, \dots, i_k)| = \begin{vmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \dots & a_{i_1 i_k} \\ a_{i_2 i_1} & a_{i_2 i_2} & \dots & a_{i_2 i_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_k i_1} & a_{i_k i_2} & \dots & a_{i_k i_k} \end{vmatrix}$$

é um menor principal de A de ordem k

O número de subdeterminantes principais de ordem k de uma matriz de ordem n é $\binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!}$, ou seja, o número de combinações sem repetição de n elementos tomados em grupos de k elementos. Portanto, o número total de menores principais é $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n - 1$.

Exemplo 11: Seja a matriz $E_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, que é obtida a partir da identidade I_3 , permutando as linhas dois e três entre si, ou que

resulta na mesma matriz E_{13} , permutando as colunas dois e três entre si. Observe que:

$$E_{23} \cdot A \cdot E_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{13} & a_{31} \\ a_{31} & a_{33} & a_{32} \\ a_{21} & a_{23} & a_{22} \end{pmatrix}.$$

Observe que $E_{23} \cdot A \cdot E_{23}$ resulta da matriz A , permutando as linhas dois e três e as colunas dois e três entre elas. Na matriz $E_{23} \cdot A \cdot E_{23}$, $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$ é o menor líder de ordem 2, que é também o menor principal $|A(1,3)|$ da matriz A . A matriz E_{23} é dita uma matriz elementar ou de permutação e observe que $E_{23} \cdot E_{23} = I$, ou seja, essa matriz é sua própria inversa. Além disso, $E_{23}' = E_{23}$.

Observe que uma matriz elementar E é uma matriz ortogonal, isto é, a sua inversa é a sua transposta: $E^{-1} = E = E'$. Digamos que $P = E_k \cdots E_2 \cdot E_1$ seja um produto de matrizes elementares, como cada uma delas é ortogonal, então o produto P é ortogonal:

$$P^{-1} = (E_k \cdots E_2 \cdot E_1)^{-1} = E_1^{-1} \cdot E_2^{-1} \cdots E_k^{-1} = E_1' \cdot E_2' \cdots E_k' = (E_k \cdots E_2 \cdot E_1)' = P'.$$

Sempre é possível, através de matrizes de permutação, colocar $|A(i_1, i_2, \dots, i_k)| = \begin{vmatrix} a_{i_1 i_1} & \cdots & a_{i_1 i_k} \\ \vdots & \ddots & \vdots \\ a_{i_k i_1} & \cdots & a_{i_k i_k} \end{vmatrix}$, $i_1 < i_2 < \dots < i_k$, que é um menor principal da matriz A , na posição de menor principal líder de ordem k , através de um produto adequado $P' \cdot A \cdot P$.

De modo mais geral, dizemos que uma matriz $P = E_k \cdots E_2 \cdot E_1$, produto de matrizes elementares, é uma matriz de permutação e que $P' \cdot A \cdot P$ é um “rearranjo principal de A ”⁴. Portanto, através de um rearranjo principal de A , sempre é possível colocar um menor principal de A numa posição de menor líder.

Exemplo 12: Sejam $E_{12} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, $E_{23} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ e $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$, então

$E_{23} \cdot E_{12} \cdot A \cdot E_{12} \cdot E_{23} = \begin{pmatrix} a_{22} & a_{23} & a_{21} & a_{24} \\ a_{32} & a_{33} & a_{31} & a_{34} \\ a_{12} & a_{13} & a_{11} & a_{14} \\ a_{42} & a_{43} & a_{41} & a_{44} \end{pmatrix}$. Na matriz $E_{23} \cdot E_{12} \cdot A \cdot E_{12} \cdot E_{23}$ o menor $|A(2,3)|$ é o líder de

ordem 2. Observe que $(E_{23} \cdot E_{12})^{-1} = E_{12}^{-1} \cdot E_{23}^{-1} = E_{12} \cdot E_{23}$. A matriz $P_{\pi}^{-1} = E_{23} \cdot E_{12} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ faz com

que as linhas 2 e 3 da matriz A passem para as duas primeiras posições; a matriz $P_{\pi} = (E_{23} \cdot E_{12})^{-1} = E_{12} \cdot E_{23} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ faz o mesmo, porém com as

colunas: $E_{23} \cdot E_{12} \cdot A \cdot E_{12} \cdot E_{23} = P_{\pi}^{-1} \cdot A \cdot P_{\pi} = P'_{\pi} \cdot A \cdot P_{\pi}$. A matriz P_{π}^{-1} resulta do processo de levar as linhas dois e três da matriz identidade I_4 para as primeiras posições; a matriz P_{π} resulta de levar as colunas dois e três da matriz identidade I_4 para as posições um e dois. Também dizemos que P_{π} P_{π}^{-1} são matrizes de permutação, ainda que não elementares, pois são produto de duas elementares.

Portanto, utilizando as operações elementares $E_{1i_1}, E_{2i_2}, \dots, E_{ki_k}$, levamos o determinante $\begin{vmatrix} a_{i_1 i_1} & \dots & a_{i_1 i_k} \\ \vdots & \ddots & \vdots \\ a_{i_k i_1} & \dots & a_{i_k i_k} \end{vmatrix}$, $i_1 < i_2 < \dots < i_k$, para a posição líder de ordem k , utilizando as operações elementares $E_{1i_1}, E_{2i_2}, \dots, E_{ki_k}$: fazemos as linhas/colunas $i_1, i_2, \dots, i_k$ da matriz A ocuparem as primeiras posições na matriz $P_{\pi}^{-1} \cdot A \cdot P_{\pi}$, em que $P_{\pi} = E_{1i_1} \dots E_{ki_k}$:

$$A^{\pi} = P_{\pi}^{-1} \cdot A \cdot P_{\pi} = (E_{ki_k} \dots E_{1i_1}) \cdot A \cdot (E_{1i_1} \dots E_{ki_k}) = \begin{pmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \dots & a_{i_1 i_k} & \dots & \dots \\ a_{i_2 i_1} & a_{i_2 i_2} & \dots & a_{i_2 i_k} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{i_k i_1} & a_{i_k i_2} & \dots & a_{i_k i_k} & \dots & \dots \\ \vdots & \vdots & \dots & \vdots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

A submatriz principal líder de ordem k de A^{π} é a submatriz principal $A(i_1, i_2, \dots, i_k)$ de A . Portanto, toda a submatriz principal de A é um menor principal líder de A^{π} .

Seja $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, então $P_\pi^{-1} \cdot x = \begin{pmatrix} x_{i_1} \\ \vdots \\ x_{i_k} \\ \vdots \end{pmatrix}$. Na verdade, o que a matriz P_π^{-1} faz é modificar o ordenamento dos eixos coordenados, passando os eixos $i_1 < i_2 < \dots < i_k$ para as primeiras posições, ou seja, estamos trocando o sistema de coordenadas $x_1, \dots, x_k, x_{k+1}, \dots, x_n$ por $x_{i_1}, \dots, x_{i_k}, \dots$.

A matriz P_π^{-1} é a matriz da troca de coordenadas de $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ para $x^\pi = \begin{pmatrix} x_{i_1} \\ \vdots \\ x_{i_k} \\ \vdots \end{pmatrix}$ e P_π é a troca de x^π para x . Dada uma forma quadrática $\mathcal{Q}(x) = x' \cdot A \cdot x$, obviamente o número $\mathcal{Q}(x)$ não se altera através de uma troca dos “nomes” dos eixos, o que era o eixo x_{i_k} passou a se chamar x_k , etc., assim $\mathcal{Q}(x) = x' \cdot A \cdot x = (P_\pi \cdot x^\pi)' \cdot A \cdot (P_\pi \cdot x^\pi) = x^{\pi'} \cdot (P_\pi' \cdot A \cdot P_\pi) \cdot x^\pi = \mathcal{Q}(x^\pi)$. Portanto, a matriz da forma no sistema de coordenadas x^π é $P_\pi' \cdot A \cdot P_\pi = A^\pi$.

Se estamos investigando a definição da forma e associando a definição à propriedades da matriz A , então essas propriedades também são satisfeitas pela matriz A^π . É importante observar que a definição de A e A^π é a mesma, pois representam a mesma forma quadrática em sistemas de coordenadas diferentes. Um sistema e outro se diferenciam apenas pelos “nomes” atribuídos aos eixos: o que é o eixo i_k num sistema é o eixo k no outro e vice-versa. A diferença entre A e A^π apenas reflete que os eixos coordenados estão ordenados de maneira diferente. Se A representa a forma no sistema $x_1, \dots, x_n$, então A^π representa a forma no sistema $x_1^\pi, \dots, x_n^\pi$.

5.2 Critério de Jacobi

Os critérios para a determinação da definição das formas quadráticas são atribuídos ao matemático alemão Carl Gustav Jacob Jacobi (1804-1851). Estes critérios resultam da possibilidade de reduzir qualquer forma quadrática a uma soma de quadrados através da chamada “Fórmula de Jacobi”⁵. A argumentação que

apresentamos na sequência é a mesma do Debreu⁶ que, por sua vez, pelo menos nesta parte do artigo, procede da maneira usual via o “Método de Jacobi”.

Seja uma troca do sistema de coordenadas $x_1, \dots, x_n$ por um sistema $\bar{x}_1, \dots, \bar{x}_n$, tal que

$$\begin{aligned}\bar{x}_1 &= L_1(x) = x_1 + l_{12} \cdot x_2 + \dots + l_{1n} \cdot x_n, \\ \bar{x}_2 &= L_2(x) = 0 \cdot x_1 + x_2 + \dots + l_{2n} \cdot x_n, \\ &\vdots \\ \bar{x}_n &= L_n(x) = 0 \cdot x_1 + 0 \cdot x_2 + \dots + x_n,\end{aligned}$$

$L = \begin{pmatrix} 1 & l_{12} & \dots & l_{1n} \\ 0 & 1 & \dots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$, em que L é uma matriz triangular superior. Inicialmente, vamos analisar um resultado auxiliar necessário. Seja $Q = x' \cdot A \cdot x$, e digamos que $|A_1| \neq 0, \dots, |A_{n-1}| \neq 0$, então existem L e $c_1, \dots, c_n$, tais que $Q = x' \cdot A \cdot x = c_1 \cdot L_1^2 + c_2 \cdot L_2^2 + \dots + c_s \cdot L_s^2 + \dots + c_n \cdot L_n^2$. Vejamos exemplos.

Exemplo 13: Seja a forma quadrática $Q(x_1, x_2) = (x_1 \ x_2) \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a \cdot x_1^2 + 2 \cdot b \cdot x_1 \cdot x_2 + c \cdot x_2^2$, tal que $a \neq 0$, sendo assim facilmente vemos que $a \cdot x_1^2 + 2 \cdot b \cdot x_1 \cdot x_2 + c \cdot x_2^2 = a \cdot \left(x_1 + \frac{b}{a} \cdot x_2\right)^2 + \left(-\frac{b^2}{a} + c\right) \cdot x_2^2$. Portanto, se escolhermos a troca de coordenadas $\bar{x}_1 = x_1 + \frac{b}{a} \cdot x_2$ e $\bar{x}_2 = x_2$, então a forma quadrática acima se expressa como uma soma de quadrados: $Q(x_1, x_2) = Q(\bar{x}_1, \bar{x}_2) = a \cdot \bar{x}_1^2 + \left(-\frac{b^2}{a} + c\right) \cdot \bar{x}_2^2$. A matriz de troca das coordenadas x_1, x_2 para $\bar{x}_1, \bar{x}_2$ é $L = \begin{pmatrix} 1 & \frac{b}{a} \\ 0 & 1 \end{pmatrix}$ e a troca de $\bar{x}_1, \bar{x}_2$ para x_1, x_2 é a inversa $L^{-1} = \begin{pmatrix} 1 & -\frac{b}{a} \\ 0 & 1 \end{pmatrix}$. Podemos concluir, desse modo, que

$$(L^{-1})' \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot L^{-1} = \begin{pmatrix} 1 & 0 \\ -\frac{b}{a} & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} 1 & -\frac{b}{a} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & c - \frac{b^2}{a} \end{pmatrix} = \begin{pmatrix} |A_1| & 0 \\ 0 & \frac{|A_2|}{|A_1|} \end{pmatrix}, \quad \text{ou} \quad \text{ainda,}$$

$$L' \cdot \begin{pmatrix} a & 0 \\ 0 & c - \frac{b^2}{a} \end{pmatrix} \cdot L = \begin{pmatrix} 1 & 0 \\ \frac{b}{a} & 1 \end{pmatrix} \cdot \begin{pmatrix} a & 0 \\ 0 & c - \frac{b^2}{a} \end{pmatrix} \cdot \begin{pmatrix} 1 & \frac{b}{a} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \text{ Observe que } a = |A_1| \neq 0 \text{ e que } -\frac{b^2}{a} + c = \frac{|A_2|}{|A_1|}.$$

Exercício 3

Verifique diretamente o cálculo abaixo

$$\begin{pmatrix} x_1 + \frac{b}{a} \cdot x_2 & x_2 \end{pmatrix} \cdot \begin{pmatrix} a & 0 \\ 0 & c - \frac{b^2}{a} \end{pmatrix} \cdot \begin{pmatrix} x_1 + \frac{b}{a} \cdot x_2 \\ x_2 \end{pmatrix} = (x_1 \ x_2) \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Portanto, a forma é positiva-definida se, e só se, $|A_1| > 0$ e $|A_2| > 0$; a forma é negativa-definida se, e só se, $|A_1| < 0$ e $|A_2| > 0$.

Exercício 4

Seja a forma quadrática

$$Q = (x_1 \ x_2 \ x_3) \cdot \begin{pmatrix} 4 & 5/2 & 3 \\ 5/2 & 1 & -3 \\ 3 & -3 & -3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 4 \cdot x^2 + y^2 - 3 \cdot z^2 + 5 \cdot x \cdot y + 6 \cdot x \cdot z - 6 \cdot y \cdot z. \text{ Observe que tal que } |A_1| = 4, |A_2| = -\frac{9}{4}$$

e que $|A_3| = -\frac{333}{4}$. Verifique diretamente que

$$Q(x_1, x_2, x_3) = \begin{pmatrix} x_1 + \frac{5}{8} \cdot x_2 + \frac{3}{4} \cdot x_3 & x_2 + \frac{26}{3} \cdot x_3 & x_3 \end{pmatrix} \cdot \begin{pmatrix} 4 & 0 & 0 \\ 0 & -\frac{9}{16} & 0 \\ 0 & 0 & 37 \end{pmatrix} \cdot \begin{pmatrix} x_1 + \frac{5}{8} \cdot x_2 + \frac{3}{4} \cdot x_3 \\ x_2 + \frac{26}{3} \cdot x_3 \\ x_3 \end{pmatrix} \text{ e que } \begin{pmatrix} 4 & 0 & 0 \\ 0 & -\frac{9}{16} & 0 \\ 0 & 0 & 37 \end{pmatrix} = \begin{pmatrix} |A_1| & 0 & 0 \\ 0 & \frac{|A_2|}{|A_1|} & 0 \\ 0 & 0 & \frac{|A_3|}{|A_2|} \end{pmatrix}. \text{ A}$$

matriz da troca das coordenadas de x_1, x_2, x_3 para $\bar{x}_1, \bar{x}_2, \bar{x}_3$ é $L = \begin{pmatrix} 1 & \frac{5}{8} & \frac{3}{4} \\ 0 & 1 & \frac{26}{3} \\ 0 & 0 & 1 \end{pmatrix}$. Qual é o

resultado do produto matricial $\begin{pmatrix} 1 & 0 & 0 \\ \frac{5}{8} & 1 & 0 \\ \frac{3}{4} & \frac{26}{3} & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 0 & 0 \\ 0 & -\frac{9}{16} & 0 \\ 0 & 0 & 37 \end{pmatrix} \cdot \begin{pmatrix} 1 & \frac{5}{8} & \frac{3}{4} \\ 0 & 1 & \frac{26}{3} \\ 0 & 0 & 1 \end{pmatrix}$? Não necessário

fazer os cálculos para obter o resultado. Observe que a forma é indefinida.

Nas duas situações acima, expressamos as formas quadráticas como estamos sugerindo.

Exemplo 14: Tratemos de maneira geral o caso da matriz $A, 3 \times 3$. Seja

$$Q(x_1, x_2, x_3) = x' \cdot A \cdot x = (x_1 \ x_2 \ x_3) \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Como, por hipótese, $a_{11} = |A_1| \neq 0$, observe primeiramente que $a_{11} \cdot \left(x_1 + \frac{a_{12}}{a_{11}} \cdot x_2 + \frac{a_{13}}{a_{11}} \cdot x_3 \right)^2 = (a_{11} \cdot x_1^2 + 2 \cdot a_{12} \cdot x_1 \cdot x_2 + 2 \cdot a_{13} \cdot x_1 \cdot x_3) + \frac{a_{12}^2}{a_{11}} \cdot x_2^2 + \frac{a_{13}^2}{a_{11}} \cdot x_3^2 + \frac{2 \cdot a_{12} \cdot a_{13}}{a_{11}} \cdot x_2 \cdot x_3$. A forma quadrática se expressa como $Q(x) =$

$$a_{11} \cdot \left(x_1 + \frac{a_{12}}{a_{11}} \cdot x_2 + \frac{a_{13}}{a_{11}} \cdot x_3 \right)^2 - \frac{a_{12}^2}{a_{11}} \cdot x_2^2 - \frac{a_{13}^2}{a_{11}} \cdot x_3^2 - \frac{2 \cdot a_{12} \cdot a_{13}}{a_{11}} \cdot x_2 \cdot x_3 + a_{22} \cdot x_2^2 + 2 \cdot a_{23} \cdot x_2 \cdot x_3 + a_{33} \cdot x_3^2.$$

Observe

que

$$-\frac{a_{12}^2}{a_{11}} \cdot x_2^2 - \frac{a_{13}^2}{a_{11}} \cdot x_3^2 - \frac{2 \cdot a_{12} \cdot a_{13}}{a_{11}} \cdot x_2 \cdot x_3 + a_{22} \cdot x_2^2 + 2 \cdot a_{23} \cdot x_2 \cdot x_3 + a_{33} \cdot x_3^2 =$$

$$\left(a_{22} - \frac{a_{12}^2}{a_{11}}\right) \cdot x_2^2 + \left(a_{33} - \frac{a_{13}^2}{a_{11}}\right) \cdot x_3^2 + 2 \cdot a_{23} \cdot x_2 \cdot x_3 = \begin{pmatrix} x_2 & x_3 \end{pmatrix} \cdot \begin{pmatrix} \frac{a_{22} \cdot a_{11} - a_{12}^2}{a_{11}} & \frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} \\ \frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} & \frac{a_{33} \cdot a_{11} - a_{13}^2}{a_{11}} \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}.$$

Observe também que

$\frac{a_{22} \cdot a_{11} - a_{12}^2}{a_{11}} = \frac{|A_2|}{|A_1|}$. Ora, pelo que vimos para o caso 2×2 , caso $\frac{|A_2|}{|A_1|} \neq 0$, isto é, $|A_2| \neq 0$, podemos expressar como

$$\begin{aligned} \begin{pmatrix} x_2 & x_3 \end{pmatrix} \cdot \begin{pmatrix} \frac{a_{22} \cdot a_{11} - a_{12}^2}{a_{11}} & \frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} \\ \frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} & \frac{a_{33} \cdot a_{11} - a_{13}^2}{a_{11}} \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \\ \frac{|A_2|}{|A_1|} \cdot \left(x_2 + \frac{\left(\frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} \right)}{\left(\frac{|A_2|}{|A_1|} \right)} \cdot x_3 \right)^2 + \left(\frac{a_{33} \cdot a_{11} - a_{13}^2}{a_{11}} - \frac{\left(\frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} \right)^2}{\left(\frac{|A_2|}{|A_1|} \right)} \right) \cdot x_3^2 = \\ \frac{|A_2|}{|A_1|} \cdot \left(x_2 + \frac{\left(\frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} \right)}{\left(\frac{|A_2|}{|A_1|} \right)} \cdot x_3 \right)^2 + \left(\frac{|A_3|}{|A_2|} \right) \cdot x_3^2, \text{ pois } \frac{a_{33} \cdot a_{11} - a_{13}^2}{a_{11}} - \frac{\left(\frac{a_{23} \cdot a_{11} - a_{12} \cdot a_{13}}{a_{11}} \right)^2}{\left(\frac{|A_2|}{|A_1|} \right)} = \frac{|A_3|}{|A_2|}. \end{aligned}$$

Conseguimos mostrar o resultado para matrizes 3×3 porque ele é válido para matrizes 2×2 . Se demonstramos o resultado para matrizes 3×3 , podemos prosseguir para demonstrá-lo para matrizes 4×4 . A única exigência é que $|A_1| \neq 0$, $|A_2| \neq 0$, $\dots, |A_{n-1}| \neq 0$. Logo, temos o seguinte lema:

Lema 1

Seja uma forma quadrática $\mathcal{Q} = x' \cdot A \cdot x$, em que A é uma matriz simétrica $n \times n$ tal que $|A_1| \neq 0$, $|A_2| \neq 0$, $\dots, |A_{n-1}| \neq 0$, então existe uma matriz triangular L como acima, tal que $\mathcal{Q} = x' \cdot A \cdot x = c_1 \cdot L_1^2 + \dots + c_i \cdot L_i^2 + \dots + c_n \cdot L_n^2$, em que $c_1 = \frac{|A_1|}{|A_0|}, \dots, c_i = \frac{|A_{i-1}|}{|A_{i-2}|}, \dots, c_n = \frac{|A_n|}{|A_{n-1}|}$. Apenas para conveniência de notação, definimos $|A_0| = 1$, portanto, $c_1 = \frac{|A_1|}{|A_0|} = |A_1|$.

Teorema 5

Seja A simétrica $n \times n$. $x' \cdot A \cdot x = \sum_{r=1}^n \frac{|A_r|}{|A_{r-1}|} \cdot L_r^2$ se, e somente se, $|A_{r-1}| \neq 0$, para $r = 1, \dots, n-1$.

$$\Rightarrow \text{óbvio.} \quad \Leftarrow \text{ Se } |A_1| \neq 0, \dots, |A_{n-1}| \neq 0, \text{ então } \mathcal{Q} = \frac{|A_1|}{|A_0|} \cdot L_1^2 + \frac{|A_2|}{|A_1|} \cdot L_2^2 + \dots + \frac{|A_n|}{|A_{n-1}|} \cdot L_n^2,$$

pelo Teorema anterior.

Teorema 6 (Critérios de Jacobi)

i) $x' \cdot A \cdot x > 0, \forall x \neq 0$ se, e somente se, $|A_r| > 0, r = 1, \dots, n$;

ii) $x' \cdot A \cdot x < 0, \forall x \neq 0$ se, e somente se, $(-1)^r \cdot |A_r| > 0, r = 1, \dots, n$.

Pelo Teorema anterior, $x' \cdot A \cdot x = \frac{|A_1|}{|A_0|} \cdot L_1^2 + \dots + \frac{|A_n|}{|A_{n-1}|} \cdot L_n^2$.

⊞ Se $|A_r| > 0, r = 1, \dots, n$, então $x' \cdot A \cdot x > 0$.

⊞ Etapa 1: $(1 \ 0 \ \dots \ 0) \cdot A \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1' \cdot A \cdot e_1 = \frac{|A_1|}{|A_0|} > 0$. Por convenção, $|A_0| = 1$. Logo $|A_1| > 0$. Etapa

2: $(0 \ 1 \ \dots \ 0) \cdot A \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = e_2' \cdot A \cdot e_2 = \frac{|A_2|}{|A_1|} > 0$, como acabamos de mostrar que $|A_1| > 0$, então $|A_2| > 0$.

E assim, prosseguimos até mostrar que $|A_n| > 0$.

Caso a forma seja negativa-definida. Etapa 1: $(1 \ 0 \ \dots \ 0) \cdot A \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1' \cdot A \cdot e_1 = \frac{|A_1|}{|A_0|} < 0$.

Por convenção, $|A_0| = 1$. Logo $|A_1| < 0$. Etapa 2: $(0 \ 1 \ \dots \ 0) \cdot A \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = e_2' \cdot A \cdot e_2 = \frac{|A_2|}{|A_1|} < 0$, como $|A_1| < 0$,

então $|A_2| > 0$. E assim prosseguimos até mostrar que $(-1)^n \cdot |A_n| > 0$. Em outras palavras, a forma quadrática é negativa-definida se os determinantes menores principais alternam o sinal.

Lema 2

i) $x' \cdot A \cdot x \geq 0$ se e somente se $x' \cdot A \cdot x + \alpha \cdot x' \cdot x > 0, \forall \alpha > 0, x \neq 0$.

ii) $x' \cdot A \cdot x \leq 0$ se e somente se $x' \cdot A \cdot x + \alpha \cdot x' \cdot x < 0, \forall \alpha < 0, x \neq 0$.

Prova: i)

⊞ $\alpha \cdot x' \cdot x = \alpha \cdot \|x\|^2 > 0$, para $\alpha > 0$, como $x' \cdot A \cdot x \geq 0$, então $x' \cdot A \cdot x + \alpha \cdot x' \cdot x > 0$, para $x \neq 0$.

⊞ se $x' \cdot A \cdot x + \alpha \cdot x' \cdot x > 0$, então $\lim_{\alpha \rightarrow 0^+} x' \cdot A \cdot x + \alpha \cdot x' \cdot x = x' \cdot A \cdot x \geq 0$.

Caso ii) análogo ao caso i).

Nos resultados a seguir, tomamos a permutação π adequada em cada caso para trazer um menor principal $|A(i_1, \dots, i_k)|$, de ordem r da

matriz A , para a posição líder na matriz A^r que indicamos por $|A_r^r|$.

Teorema 7 (Critérios de Jacobi)

iii) $x' \cdot A \cdot x \geq 0$ se, e somente se, $|A_r^r| \geq 0, \forall r=1, \dots, n, \forall \pi$, permutação;

iv) $x' \cdot A \cdot x \leq 0$ se, e somente se, $(-1)^r \cdot |A_r^r| \geq 0, \forall r=1, \dots, n, \forall \pi$, permutação.

Prova:

Caso i) $\Rightarrow$ Como $x^{\pi'} \cdot A^{\pi} \cdot x^{\pi} \geq 0$, desse modo, para $\alpha > 0$ e $x \neq 0$, $x^{\pi'} \cdot A^{\pi} \cdot x^{\pi} + x^{\pi'} \cdot (\alpha \cdot I) \cdot x^{\pi} = x^{\pi'} \cdot (A^{\pi} + \alpha \cdot I) \cdot x^{\pi} > 0$, ou seja, a forma quadrática dada pela matriz $A^{\pi} + \alpha \cdot I$ é positiva-definida. Logo, $|A_r^{\pi} + \alpha \cdot I_r| > 0$. Portanto, $|A_r^{\pi}| = \lim_{\alpha \rightarrow 0^+} |A_r^{\pi} + \alpha \cdot I_r| \geq 0$.

$\Leftarrow$ $|A_r + \alpha \cdot I| = \alpha^r + \sum_{i=0}^{r-1} S_{r-i} \alpha^i > 0, \forall \alpha > 0$, e $r=1, \dots, n$, pois $S_{r-i} \geq 0$, para $r=1, \dots, n$. Então, $(A + \alpha \cdot I)$ é positiva-definida, para $\forall \alpha > 0$. Portanto, A é positiva-semidefinida.

Caso ii) $\Rightarrow$ Como $x' \cdot A^{\pi} \cdot x \leq 0$, desse modo, para $\alpha < 0$ e $x \neq 0$, $x' \cdot A^{\pi} \cdot x + x' \cdot (\alpha \cdot I) \cdot x < 0$. Logo, $(-1)^r \cdot |A_r^{\pi} + \alpha \cdot I_r| > 0$. Portanto, $(-1)^r \cdot |A_r^{\pi}| = \lim_{\alpha \rightarrow 0^+} (-1)^r \cdot |A_r^{\pi} + \alpha \cdot I_r| \geq 0$.

$\Leftarrow$ $(-1)^r \cdot |A_r + \alpha \cdot I| = (-1)^r \cdot \alpha^r + \sum_{i=0}^{r-1} (-1)^r \cdot S_{r-i} \cdot \alpha^i = (-\alpha)^r + \sum_{i=0}^{r-1} (-1)^{r-i} \cdot S_{r-i} \cdot (-\alpha)^i > 0$, para todo $\alpha < 0$ e $r=1, \dots, n$. Então, $(A + \alpha \cdot I)$ é negativa-definida, para $\forall \alpha < 0$. Portanto, A é negativa-semidefinida.

Reunindo os teoremas anteriores, repetimos os resultados obtidos

Teorema 8 (Critérios de Jacobi)

i) $x' \cdot A \cdot x > 0, \forall x \neq 0$ se, e somente se, $|A_r| > 0, r=1, \dots, n$;

ii) $x' \cdot A \cdot x < 0, \forall x \neq 0$ se, e somente se, $(-1)^r \cdot |A_r| > 0, r=1, \dots, n$;

iii) $x' \cdot A \cdot x \geq 0$ se, e somente se, $|A_r^{\pi}| \geq 0, \forall r=1, \dots, n, \forall \pi$;

iv) $x' \cdot A \cdot x \leq 0$ se, e somente se, $(-1)^r \cdot |A_r^{\pi}| \geq 0, \forall r=1, \dots, n, \forall \pi$;

iv) $x' \cdot A \cdot x$ é indefinida se, e somente se, não satisfaz os critérios iii) e iv).

O item i) afirma que a condição necessária e suficiente para A ser positiva-definida é que todos os menores principais líderes de A sejam positivos; para A ser negativa-definida, é que os menores

principais líderes de ordem par sejam positivos e que os de ordem ímpar sejam negativos; para A ser positiva-semidefinida, é que todos os menores principais sejam maiores ou iguais a zero; para A ser negativa-semidefinida, é que os menores principais de ordem par sejam não-negativos e que os de ordem ímpar sejam não-positivos. Se algum menor principal de ordem par for negativo ou algum menor principal de ordem ímpar for positivo, então a forma é não-definida.

Suponhamos que $S \subseteq \mathbb{R}^n$ é um subconjunto convexo.

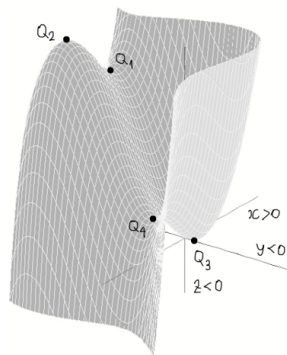
Teorema 9

- i) $f: S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ é **convexa** se, e somente se, $|\nabla^2 f(x)|_r \geq 0, \forall r, r=1, \dots, n$ em S ;
- ii) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ é **côncava** se, e somente se, $(-1)^r \cdot |\nabla^2 f(x)|_r \geq 0, \forall r$, em S ;
- iii) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, se $|\nabla^2 f(x)|_r > 0, r=1, \dots, n$, em S então f é **estritamente convexa** em S ;
- iv) $f: I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, se $(-1)^r \cdot |\nabla^2 f(x)|_r > 0, r=1, \dots, n$ em S , então f é **estritamente côncava** em S .

Teorema 10

- i) $\nabla f(x^*) = f'(x^*) = 0$ e $|\nabla^2 f(x^*)|_r > 0, r=1, \dots, n$, então x^* é um **mínimo local**;
- ii) $\nabla f(x^*) = f'(x^*) = 0$ e $(-1)^r \cdot |\nabla^2 f(x^*)|_r > 0, r=1, \dots, n$, então x^* é um **máximo local**.

Exemplo 15: Seja a função $f(x, y) = x^3 - y^3 + x^2 + 3 \cdot y^2 - x$. Essa função tem quatro pontos críticos: $P_1(-1, 0), P_2(-1, 2), P_3(1/3, 0)$ e $P_4(1/3, 2)$ que são uma sela, um máximo local, um mínimo local e uma sela, respectivamente. Veja o desenho abaixo. Os pontos $Q_1(-1, 0, 1), Q_2(-1, 2, 5), Q_3(1/3, 0, -5/27)$ e $Q_4(1/3, 2, 103/27)$ estão marcados:



Como $f''(x,y) = \begin{pmatrix} 6 \cdot x + 2 & 0 \\ 0 & -6 \cdot y + 6 \end{pmatrix}$, $f''(-1,0) = \begin{pmatrix} -4 & 0 \\ 0 & 6 \end{pmatrix}$, $f''(-1,2) = \begin{pmatrix} -4 & 0 \\ 0 & -6 \end{pmatrix}$, $f''(1/3,0) = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}$ e $f''(1/3,2) = \begin{pmatrix} 4 & 0 \\ 0 & -6 \end{pmatrix}$ são formas indefinida, negativa-definida, positiva-definida e indefinida, respectivamente.

Exercício 5

Exercício 5.1: Classifique os pontos críticos quanto a mínimo, máximo local ou sela das funções abaixo:

- i) $f(x,y) = -x^2 \cdot y + x \cdot y^2 + x \cdot y$;
- ii) $f(x,y) = -x^2 \cdot y - x \cdot y^2 + x \cdot y$;
- iii) $f(x,y) = x^3 + 2 \cdot x \cdot y - 3 \cdot x^2 + y^2 + 1$;
- iv) $f(x,y,z) = x^2 + 2 \cdot x \cdot y + y^3 + z^2 - 5 \cdot y + 4 \cdot z$
- v) $f(x,y,z) = x^2 + 2 \cdot y^2 - 3 \cdot z^2 - x - 2 \cdot y + 3 \cdot z$;
- vi) $f(x,y,z) = x^3 + 2 \cdot x \cdot y + y^2 + z^2 - 5 \cdot x - 4 \cdot z$;
- vii) $f(x,y,z) = x^3 + y^3 + z^3 - 3 \cdot x - 3 \cdot y - 3 \cdot z + 2$;
- viii) $f(x,y,z) = x^3 + 2 \cdot x \cdot y + y^2 + z^2 - 5 \cdot x - 4 \cdot z$;
- ix) $f(x,y,z) = -x^3 + 2 \cdot x \cdot z + 3 \cdot y - y^2 - z^2$;
- x) $f(x,y,z) = \frac{x^4}{4} + x \cdot y + \frac{y^2}{2} + z^2$.

Exercício 5.2: Classifique as funções quanto à concavidade ou convexidade:

- i) $f(x,y,z) = -x \cdot \ln(x) - y^2 - z \cdot \ln(z)$, $x > 0, y > 0, z > 0$;
- ii) $f(x,y,z) = -x^2 - 2 \cdot y^2 - z^2 + x \cdot y + z$;

$$\text{iii)} \quad f(x, y) = y + x^2 + y^3;$$

$$\text{iv)} \quad f(x, y, z) = x^2 + 2 \cdot y^2 - 3 \cdot z^2 - x - 2 \cdot y + 3 \cdot z;$$

$$\text{v)} \quad f(x, y, z) = -x^2 - y^2 - z^2 + x \cdot y + x + y;$$

$$\text{vi)} \quad f(x, y, z) = -x^2 - 2 \cdot y^2 - 4 \cdot z^2 + 2 \cdot x \cdot z - 4 \cdot y \cdot z - 2 \cdot x - 6 \cdot z;$$

vii) Estabeleça condições sobre a , b e c de maneira que

$$f(x, y, z) = \frac{a}{2} \cdot x^2 + c \cdot x \cdot y + \frac{b}{2} \cdot y^2 + \frac{c}{2} \cdot z^2 \text{ seja estritamente côncava;}$$

$$\text{viii)} \quad f(x, y) = 6 \cdot x^{\frac{1}{3}} \cdot y^{\frac{1}{3}} - 2 \cdot x - 2 \cdot y, \quad x > 0, y > 0;$$

ix) Mostre que $f(x, y) = x^a \cdot y^{1-a}$ para $x, y > 0$ é côncava se, e só se, $0 \leq a \leq 1$.

6. Formas quadráticas restritas a subespaços

Na maioria dos nossos problemas de otimização, a função objetivo f está restrita a domínios $S \subseteq \mathbb{R}^n$ definidos por superfícies de dimensão menores que n . Isso implica que a definição da derivada f' não precisa ser determinada em todo o espaço, mas apenas nas direções viáveis determinadas pelas restrições.

Seja o problema:

$$\begin{aligned} \min \quad & f(x_1, x_2, \dots, x_n) \\ \text{sa} \quad & g_1(x_1, x_2, \dots, x_n) = c_1 \\ & \vdots \\ & g_m(x_1, x_2, \dots, x_n) = c_m. \end{aligned}$$

Digamos que $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ seja um ponto que satisfaz as restrições e o ponto crítico do lagrangeano $\mathcal{L}(x) = f(x) - \lambda_1 \cdot (g_1(x) - c_1) - \dots - \lambda_m \cdot (g_m(x) - c_m)$.

Suponhamos que $\mathcal{L}''(x^*)$ seja positiva-definida no subespaço dos vetores $v \in \mathbb{R}^n$ determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, isto é, dos vetores ortogonais aos gradientes $\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_m(x^*)$. Sendo assim, x^* é um mínimo local para f dadas as restrições $g_1 = c_1, g_2 = c_2, \dots, g_m = c_m$. Vejamos: $\mathcal{L}(x) = \mathcal{L}(x^*) + \nabla \mathcal{L}(x^*) \cdot (x - x^*) + \frac{1}{2} \cdot (x - x^*)' \cdot \mathcal{L}''(\bar{x}) \cdot (x - x^*)$, para algum $\bar{x}$ no segmento de reta unindo x^* e x . Observe que $\bar{x}$ não está necessariamente satisfazendo às restrições $g_j(x) = c_j, j = 1, \dots, m$. Como x^* satisfaz as condições de primeira ordem do lagrangeano, então $\nabla \mathcal{L}(x^*) = 0$. Além disso, como x^* e x

satisfazem as restrições, então $\mathcal{L}(x^*) = f(x^*)$ e $\mathcal{L}(x) = f(x)$. Portanto, $f(x) - f(x^*) = \frac{1}{2} \cdot (x - x^*)' \cdot \mathcal{L}''(\bar{x}) \cdot (x - x^*)$.

Por outro lado, para cada x , se x^* e x satisfazem a restrição, então $0 = g_j(x) - g_j(x^*) = \nabla g_j(x^*) \cdot (x - x^*) + \frac{1}{2} \cdot (x - x^*)' \cdot g_j''(\bar{x}_j) \cdot (x - x^*)$ para algum $\bar{x}_j$ entre x^* e x . Observe que $\bar{x}_j$, que depende de x^* e x , varia quando x varia. Portanto, $\frac{\nabla g_j(x^*) \cdot (x - x^*)}{\|x - x^*\|} = -\frac{1}{2} \cdot \frac{(x - x^*)' \cdot g_j''(\bar{x}_j) \cdot (x - x^*)}{\|x - x^*\|}$, como $\lim_{x \rightarrow x^*} \frac{(x - x^*)' \cdot g_j''(\bar{x}_j) \cdot (x - x^*)}{\|x - x^*\|} = 0$, então $\lim_{x \rightarrow x^*} \frac{\nabla g_j(x^*) \cdot (x - x^*)}{\|x - x^*\|} = 0$.

Digamos que exista uma sequência de pontos x_k , tal que $x_k \rightarrow x^*$, $g_1(x_k) = c_1$, $g_2(x_k) = c_2, \dots, g_m(x_k) = c_m$ e $f(x_k) \leq f(x^*)$. A sequência ou uma subsequência retirada dela, $v_k = \frac{x_k - x^*}{\|x_k - x^*\|} \in S^{n-1}$, converge para algum $v^* = \frac{x_k - x^*}{\|x_k - x^*\|} \rightarrow v^* \in S^{n-1}$. Pelo que vimos acima, $\nabla g_j(x^*) \cdot v^* = \lim_{k \rightarrow +\infty} \nabla g_j(x^*) \cdot v_k = \lim_{k \rightarrow +\infty} \frac{\nabla g_j(x^*) \cdot (x_k - x^*)}{\|x_k - x^*\|} = 0$, portanto,

$$\lim_{k \rightarrow +\infty} \frac{f(x_k) - f(x^*)}{\|x_k - x^*\|^2} = \lim_{k \rightarrow +\infty} \frac{1}{2} \cdot \frac{(x_k - x^*)'}{\|x_k - x^*\|} \cdot \mathcal{L}''(\bar{x}) \cdot \frac{(x_k - x^*)}{\|x_k - x^*\|} =$$

$\lim_{k \rightarrow +\infty} \frac{1}{2} \cdot v_k' \cdot \mathcal{L}''(\bar{x}) \cdot v_k = \frac{1}{2} \cdot v^{*'} \cdot \mathcal{L}''(x^*) \cdot v^* > 0$, pois $\mathcal{L}''(x^*)$ é positiva-definida em $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, por hipótese. Por conseguinte, $\lim_{k \rightarrow +\infty} \frac{f(x_k) - f(x^*)}{\|x_k - x^*\|^2} > 0$, mas isso é uma contradição, pois supusemos que $f(x_k) \leq f(x^*)$, o que implica $\lim_{k \rightarrow +\infty} \frac{f(x_k) - f(x^*)}{\|x_k - x^*\|^2} \leq 0$. Portanto, não existe sequência de pontos x_k tal que $x_k \rightarrow x^*$ e $f(x_k) \leq f(x^*)$. Desse modo, existe uma vizinhança de x^* , tal que os pontos x nessa vizinhança satisfazem às restrições e são da mesma maneira que $f(x) > f(x^*)$. Logo, existe uma vizinhança de x^* em que x^* é mínimo e é único.

Façamos outra demonstração. Considere uma curva $\varphi(t)$, definida num intervalo $[0, \varepsilon)$, tal que $g_1(\varphi(t)) = c_1, g_2(\varphi(t)) = c_2, \dots, g_m(\varphi(t)) = c_m$, $\varphi(0) = x^*$ e $\varphi'(0) = v \neq 0$, então $\nabla g_1(x^*) \cdot \varphi'(0) = 0, \nabla g_2(x^*) \cdot \varphi'(0) = 0, \dots, \nabla g_m(x^*) \cdot \varphi'(0) = 0$. Como $\frac{f(\varphi(t)) - f(x^*)}{t^2} = \frac{1}{2} \cdot \left(\frac{\varphi(t) - x^*}{t} \right)' \cdot \mathcal{L}''(\bar{x}) \cdot \left(\frac{\varphi(t) - x^*}{t} \right)$ para algum $\bar{x}$ entre $\varphi(t)$ e x^* , fazendo $t \rightarrow 0^+$, obtemos $\lim_{t \rightarrow 0^+} \frac{f(\varphi(t)) - f(x^*)}{t^2} = \lim_{t \rightarrow 0^+} \frac{1}{2} \cdot \left(\frac{\varphi(t) - x^*}{t} \right)' \cdot \mathcal{L}''(\bar{x}) \cdot \left(\frac{\varphi(t) - x^*}{t} \right) = \frac{1}{2} \cdot v' \cdot \mathcal{L}''(x^*) \cdot v > 0$. Portanto, numa vizinhança de x^* , $f(x) > f(x^*)$.

De maneira similar, se considerarmos o problema:

$$\begin{aligned} \max \quad & f(x_1, x_2, \dots, x_n) \\ \text{sa} \quad & g_1(x_1, x_2, \dots, x_n) = c_1 \\ & \vdots \\ & g_m(x_1, x_2, \dots, x_n) = c_m, \end{aligned}$$

x^* é um ponto que satisfaz as restrições, ponto crítico do lagrangeano $\mathcal{L}(x) = f(x) - \lambda_1 \cdot (g_1(x) - c_1) - \dots - \lambda_m \cdot (g_m(x) - c_m)$, e $\mathcal{L}''(x^*)$ é negativa-definida no subespaço dos vetores $v \in \mathbb{R}^n$ determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, isto é, o subespaço dos vetores ortogonais aos gradientes $\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_m(x^*)$, então x^* é um máximo local isolado para f , dadas as restrições $g_1 = c_1, g_2 = c_2, \dots, g_m = c_m$. Deixamos a demonstração como exercício.

Nas proposições seguintes, precisamos um resultado que está implícito desde a Lição Quatro: se uma superfície é definida pela interseção das superfícies $g_1 = c_1, \dots, g_m = c_m$, $g_1(x^*) = c_1, \dots, g_m(x^*) = c_m$ e v é um vetor tal que $\nabla g_1(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então existe uma curva $\varphi(t)$ definida num intervalo $(-\varepsilon, \varepsilon)$, tal que $\varphi(0) = x^*$ e $\varphi'(0) = v$. Em outras palavras, todo o vetor v , tal que $\nabla g_1(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, é vetor tangente à superfície determinada por $g_1 = c_1, \dots, g_m = c_m$ em x^* . A demonstração dessa proposição reservamos para a próxima lição.

Seja um vetor v tal que $\nabla g_1(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então existe uma curva $\varphi(t)$ definida num intervalo $[0, \varepsilon)$, tal que $g_1(\varphi(t)) = c_1, \dots, g_m(\varphi(t)) = c_m$, $\varphi(0) = x^*$ e $\varphi'(0) = v$. Digamos que x^* seja um mínimo local de f , sujeita às restrições $g_1 = c_1, \dots, g_m = c_m$, então

$$0 \leq \frac{f(\varphi(t)) - f(\varphi(0))}{t^2} = \frac{1}{2} \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right)' \cdot \mathcal{L}''(\bar{x}) \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right). \quad \text{Portanto,}$$

$\lim_{t \rightarrow 0} \left(\frac{\varphi(t) - \varphi(0)}{t} \right)' \cdot \mathcal{L}''(\bar{x}) \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right) = v' \cdot \mathcal{L}''(x^*) \cdot v \geq 0$. Como observamos anteriormente, $v' \cdot \mathcal{L}''(x^*) \cdot v \geq 0$ não implica que $\mathcal{L}''(x^*)$ seja positiva-definida ou positiva-semidefinida. Pode ocorrer que $\mathcal{L}''(x^*)$ seja nula.

Exemplo 16: Consideremos a função $f(x, y) = x^4 + 2 \cdot x^2 \cdot y^2 + y^4 = (x^2 + y^2)^2$ restrita à reta $x - y = 0$. Obviamente, o ponto $(0, 0)$ é um mínimo global único e $\mathcal{L}''(0, 0) = 0$.

Entretanto, se x^* é um mínimo local de f sujeita às restrições $g_1 = c_1, \dots, g_m = c_m$ e $\mathcal{L}''(x^*)$ não é nula então $\mathcal{L}''(x^*)$ é positiva-definida ou positiva-

semidefinida, isto é, $v' \cdot \mathcal{L}''(x^*) \cdot v \geq 0$, para todo o vetor v tangente à superfície determinada pelas restrições.

Similarmente, se x^* é um máximo local de f , sujeita às restrições $g_1 = c_1, \dots, g_m = c_m$ e $\mathcal{L}''(x^*)$, não é nula, então $\mathcal{L}''(x^*)$ é negativa-definida ou negativa-semidefinida, isto é, $v' \cdot \mathcal{L}''(x^*) \cdot v \leq 0$ para todo o vetor v tangente à superfície determinada pelas restrições. Esses resultados são empregados na próxima lição.

Se $\mathcal{L}''(x^*)$ é indefinida no subespaço dos vetores $v \in \mathbb{R}^n$, determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então existe $v \in \mathbb{R}^n$, ortogonal aos gradientes $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$, tal que $v' \cdot \nabla \mathcal{L}''(x^*) \cdot v > 0$. Seja uma curva $\varphi(t)$ definida num intervalo $[0, \varepsilon)$, tal que $g_1(\varphi(t)) = c_1, \dots, g_m(\varphi(t)) = c_m$, $\varphi(0) = x^*$ e $\varphi'(0) = v$, então $\lim_{t \rightarrow 0} \frac{f(\varphi(t)) - f(\varphi(0))}{t^2} = \frac{1}{2} \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right)' \cdot \mathcal{L}''(\bar{x}) \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right) = \frac{1}{2} \cdot v' \cdot \mathcal{L}''(\bar{x}) \cdot v > 0$, portanto, na proximidade de x^* , existem pontos $\varphi(t)$ satisfazendo às restrições e tais que $f(\varphi(t)) > f(x^*)$. De maneira análoga, existe $v \in \mathbb{R}^n$, determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, de modo que existe $v \in \mathbb{R}^n$, ortogonal aos gradientes $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$, tal que $v' \cdot \nabla \mathcal{L}''(x^*) \cdot v < 0$, o que implica que, na proximidade de x^* , existem pontos $\varphi(t)$ satisfazendo às restrições e tais que $f(\varphi(t)) < f(x^*)$. Em resumo, se $\mathcal{L}''(x^*)$ é indefinida no subespaço dos vetores $v \in \mathbb{R}^n$, determinado por $\nabla g_1(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então x^* não é nem mínimo nem máximo local.

Teorema 11

Dado o problema:

$$\begin{aligned} \min/\max \quad & f(x_1, x_2, \dots, x_n) \\ \text{sa} \quad & g_1(x_1, x_2, \dots, x_n) = c_1 \\ & \vdots \\ & g_m(x_1, x_2, \dots, x_n) = c_m. \end{aligned}$$

i) Digamos que $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ seja um ponto que satisfaz as restrições e seja ponto crítico do lagrangeano $\mathcal{L}(x) = f(x) - \lambda_1 \cdot (g_1(x) - c_1) - \dots - \lambda_m \cdot (g_m(x) - c_m)$. Então,

ii) se $\mathcal{L}''(x^*)$ é positiva-definida no subespaço dos vetores $v \in \mathbb{R}^n$, determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então x^* é um mínimo

local isolado para f numa vizinhança de x^* ;

iii) se $\mathcal{L}''(x^*)$ é negativa-definida no subespaço do vetores $v \in \mathbb{R}^n$, determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então x^* é um máximo local isolado para f numa vizinhança de x^* .

iv) Digamos que $\mathcal{L}''(x^*)$ é indefinida no subespaço do vetores $v \in \mathbb{R}^n$, determinado por $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então x^* não é nem mínimo e nem é máximo.

v) Se x^* é um mínimo local de f e $\mathcal{L}''(x^*) \neq 0$, então $\mathcal{L}''(x^*)$ é positiva-definida ou positiva-semidefinida, isto é, $v' \cdot \mathcal{L}''(x^*) \cdot v \geq 0$, para todo o vetor v tangente à superfície determinada pelas restrições.

vi) se x^* é um máximo local de f e $\mathcal{L}''(x^*) \neq 0$, então $\mathcal{L}''(x^*)$ é negativa-definida ou negativa-semidefinida, isto é, $v' \cdot \mathcal{L}''(x^*) \cdot v \leq 0$, para todo o vetor v tangente à superfície determinada pelas restrições.

Exemplo 17: Na parte i) do teorema acima, não podemos substituir “positiva-definida” por “positiva-semidefinida”, pois $f(x, y) = x^2 + y^3$ é tal que $f''(0, 0)$ é positiva-semidefinida e $(0, 0)$ não é mínimo local de f . De maneira similar, não podemos substituir “negativa-definida” por “negativa-semidefinida” na parte ii).

Exemplo 18: Consideremos um problema tratado anteriormente:

$$\begin{aligned} \min \quad & x^2 + y^2 + (z-1)^2 \\ \text{sa} \quad & x^2 - y^2 - z = 0, \end{aligned}$$

Isto é, o problema da determinação do ponto sobre o parabolóide $g(x, y, z) = x^2 - y^2 - z = 0$, que está mais próximo do ponto $P(0, 0, 1)$. O lagrangeano desse problema é $\mathcal{L}(x, y, z) = x^2 + y^2 + (z-1)^2 - \lambda \cdot (x^2 - y^2 - z)$. Existem três pontos críticos para o lagrangeano: $P\left(\pm \frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$, $\lambda^* = 1$, e $P(0, 0, 0)$, $\lambda^* = 2$. Sabemos que as soluções estão em $P\left(\pm \frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$. O gradiente de $\nabla g(x, y, z) = \begin{pmatrix} 2 \cdot x \\ -2 \cdot y \\ -1 \end{pmatrix}$, $\nabla g\left(\pm \frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right) = \begin{pmatrix} \pm \sqrt{2} \\ 0 \\ -1 \end{pmatrix}$, portanto, o

subespaço dos vetores ortogonais aos vetores $\nabla g\left(\pm\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ é formado pelos $v = t \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = t \cdot e_2, t \in \mathbb{R}$. A derivada segunda é $\mathcal{L}''(x) = \begin{pmatrix} 2 \cdot (1-\lambda) & 0 & 0 \\ 0 & 2 \cdot (1+\lambda) & 0 \\ 0 & 0 & 2 \end{pmatrix}$. $\mathcal{L}''\left(\pm\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, que é uma forma positiva-semidefinida em $\mathbb{R}^3$, mas no subespaço $W = \{v = t \cdot e_2 \mid t \in \mathbb{R}\}$, $v \cdot \mathcal{L}''\left(\pm\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right) \cdot v = 4 \cdot t^2$ é positiva-definida. Logo, os pontos $P\left(\pm\frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ são mínimos locais. Sabemos que são mínimos globais na restrição.

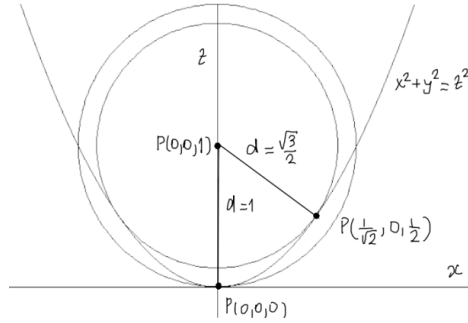
Por outro lado, sabemos que $P(0,0,0)$ não é um ponto nem de mínimo nem de máximo: $\nabla g(0,0,0) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$, e o subespaço dos vetores ortogonais ao $\nabla g(0,0,0)$ é $\mathcal{O}_{\mathbb{R}^2} = \left\{ v = \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix} \mid v_1, v_2 \in \mathbb{R} \right\}$; $\mathcal{L}''(0,0,0) = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, que é uma forma não-definida em $\mathbb{R}^3$ e, particularmente, não-definida em $\mathbb{R}^2 = \left\{ v = \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix} \mid v_1, v_2 \in \mathbb{R} \right\}$.

Exemplo 19: Seja o problema da determinação do ponto sobre o paraboloide

$g(x, y, z) = x^2 + y^2 - z = 0$, que está mais próximo do ponto $P(0,0,1)$:

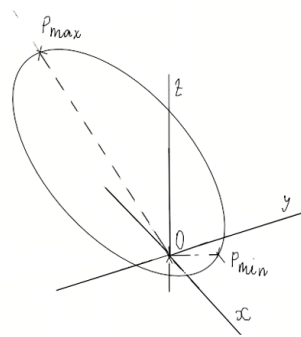
$$\begin{array}{ll} \min & x^2 + y^2 + (z-1)^2 \\ \text{sa} & x^2 + y^2 - z = 0. \end{array}$$

A derivada segunda do lagrangeano $\mathcal{L}(x, y, z) = x^2 + y^2 + (z-1)^2 - \lambda \cdot (x^2 + y^2 - z)$ é $\mathcal{L}''(x, y, z) = \begin{pmatrix} 2 \cdot (1-\lambda) & 0 & 0 \\ 0 & 2 \cdot (1-\lambda) & 0 \\ 0 & 0 & 2 \end{pmatrix}$ e $\nabla g(x, y, z) = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ -1 \end{pmatrix}$. O ponto $P(0,0,0)$, $\lambda^* = 2$, é ponto crítico do lagrangeano. $\nabla g(0,0,0) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$; o subespaço dos vetores ortogonais ao $\nabla g(0,0,0)$ é conjunto de vetores $v = \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix}$. Nesse subespaço, $\mathcal{L}''(0,0,0) = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ é negativa-definida. Logo, $P(0,0,0)$ é máximo local. Veja o desenho:



Os pontos de mínimo estão sobre a circunferência $x^2 + y^2 = \frac{1}{2}, z = \frac{1}{2}$, $\lambda^* = 1$.
 Nesses pontos, $\mathcal{L}''(x, y, z) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Nos planos gerados pelos vetores $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$,
 ortogonais aos $\nabla g(x, y, z) = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ -1 \end{pmatrix}$, $v' \cdot \mathcal{L}'' \cdot v = 2 \cdot v_3^2$. Portanto, as $\mathcal{L}''(x, y, z)$ são positivas-
 semidefinidas. Se as $\mathcal{L}''(x, y, z)$ fossem positivas-definidas, então cada
 ponto sobre a circunferência $x^2 + y^2 = \frac{1}{2}, z = \frac{1}{2}$ seria um mínimo isolado, o que
 não é o caso.

Exemplo 20: Seja o problema da determinação dos pontos sobre a
 circunferência determinada pela interseção do paraboloide
 $g_1(x, y, z) = x^2 + y^2 - 1 - z = 0$ e o plano $g_2(x, y, z) = x + y + z - 1 = 0$, que está mais próximo e mais
 distante da origem $P(0, 0, 0)$. O ponto mais próximo é $P_{\min} = P\left(\frac{1}{2} \cdot (\sqrt{7} - 1), \frac{1}{2} \cdot (\sqrt{7} - 1), 3 - \sqrt{7}\right)$
 e o ponto mais distante é $P_{\max} = P\left(-\frac{1}{2} \cdot (\sqrt{7} + 1), -\frac{1}{2} \cdot (\sqrt{7} + 1), 3 + \sqrt{7}\right)$.



Este problema pode ser formulado como:

$$\begin{array}{ll} \min/\max & x^2 + y^2 + z^2 \\ \text{sa} & x^2 + y^2 - 1 - z = 0 \\ & x + y + z - 1 = 0 \end{array}$$

A derivada segunda do lagrangeano $\mathcal{L}(x, y, z) = x^2 + y^2 + z^2 - \lambda_1 \cdot (x^2 + y^2 - 1 - z) - \lambda_2 \cdot (x + y + z - 1)$ é

$$\mathcal{L}''(x, y, z) = \begin{pmatrix} 2 \cdot (1 - \lambda) & 0 & 0 \\ 0 & 2 \cdot (1 - \lambda) & 0 \\ 0 & 0 & 2 \end{pmatrix}. \quad \text{Para o ponto } p_{\min} = P\left(\frac{1}{2} \cdot (\sqrt{7} - 1), \frac{1}{2} \cdot (\sqrt{7} - 1), 3 - \sqrt{7}\right), \quad \text{os}$$

multiplicadores são $\lambda_1 = 3 - \sqrt{7}$ e $\lambda_2 = 9 - 3 \cdot \sqrt{7}$, resultando em $\mathcal{L}''(p_{\min}) = \begin{pmatrix} 2 \cdot \sqrt{7} - 4 & 0 & 0 \\ 0 & 2 \cdot \sqrt{7} - 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

Essa forma é positiva-definida em todo $\mathbb{R}^3$ e, em particular, no subespaço ortogonal aos gradientes $\nabla g_1 = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ -1 \end{pmatrix}$ e $\nabla g_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ no ponto $p_{\min}$, que é a reta $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \cdot t, t \in \mathbb{R}$.

Para o ponto $p_{\max} = P\left(-\frac{1}{2} \cdot (\sqrt{7} + 1), -\frac{1}{2} \cdot (\sqrt{7} + 1), 3 + \sqrt{7}\right)$, os multiplicadores são $\lambda_1 = 3 + \sqrt{7}$ e $\lambda_2 = 9 + 3 \cdot \sqrt{7}$, resultando em $\mathcal{L}''(p_{\max}) = \begin{pmatrix} -2 \cdot \sqrt{7} - 4 & 0 & 0 \\ 0 & -2 \cdot \sqrt{7} - 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Essa forma é indefinida em todo $\mathbb{R}^3$, mas ela é negativa-definida no subespaço ortogonal aos gradientes $\nabla g_1 = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ -1 \end{pmatrix}$ e $\nabla g_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ no ponto $p_{\max}$, que é novamente a reta $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \cdot t, t \in \mathbb{R}$.

Portanto, podemos afirmar que $p_{\min}$ e $p_{\max}$ são pontos de mínimo e máximos locais, respectivamente.

6.1. A demonstração de Väliäho para o Teorema de Mann-Debreu

Uma “*Technical Note*”⁷, publicada em 1982, de autoria do matemático finlandês Hannu Väliäho, simplificou consideravelmente a demonstração clássica do Teorema do Debreu. Väliäho utilizou a fórmula $\left| \frac{A}{C} \middle| \frac{B}{D} \right| = |A| \cdot |D - C \cdot A^{-1} \cdot B|$, conhecida como “Fórmula de Schur”⁸.

Väliäho mostrou que a definição da forma $x' \cdot A \cdot x$, A matriz $n \times n$, no subespaço $B \cdot x = 0$, B matriz $m \times n$ de posto m , $m < n$, é a mesma definição de certa forma $y' \cdot D \cdot y$, definida em $\mathbb{R}^{n-m}$. A mesma argumentação aplica-se tanto para as formas definidas quanto para as semidefinidas.

Teorema 12 (O Complemento de Schur)

Seja A uma matriz inversível $m \times m$, então $\begin{pmatrix} A_{m \times m} & B_{m \times p} \\ C_{p \times m} & D_{p \times p} \end{pmatrix}$ é inversível. B , C e D são matrizes $m \times p$, $p \times m$ e $p \times p$, respectivamente;

$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1} \cdot B \cdot (D - C \cdot A^{-1} \cdot B)^{-1} \cdot C \cdot A^{-1} & A^{-1} \cdot B \cdot (D - C \cdot A^{-1} \cdot B)^{-1} \\ -(D - C \cdot A^{-1} \cdot B)^{-1} \cdot C \cdot A^{-1} & (D - C \cdot A^{-1} \cdot B)^{-1} \end{pmatrix}$; $\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \cdot |(D - C \cdot A^{-1} \cdot B)|$. A matriz $(D - C \cdot A^{-1} \cdot B)$ é chamada o “complemento de Schur da matriz A ”⁹.

Prova: Consideremos o sistema $A_{m \times m} \cdot X_{m \times m} + B_{m \times p} \cdot Y_{p \times m} = 0_{m \times p}$, $C_{p \times m} \cdot X_{m \times m} + D_{p \times p} \cdot Y_{p \times m} = I_{p \times p}$, X , Y e 0 são matrizes $m \times p$, $p \times p$ e $m \times p$, respectivamente. $I_{p \times p}$ é matriz identidade $p \times p$. $A \cdot X + B \cdot Y = 0$ implica que $X = -A^{-1} \cdot B \cdot Y$. Substituindo, na segunda equação, resulta $C \cdot X + D \cdot Y = -C \cdot A^{-1} \cdot B \cdot Y + D \cdot Y = I$, portanto, $(-C \cdot A^{-1} \cdot B + D) \cdot Y = I$. Logo Y é inversível e a sua inversa é $Y = (-C \cdot A^{-1} \cdot B + D)^{-1}$ e $X = -A^{-1} \cdot B \cdot (-C \cdot A^{-1} \cdot B + D)^{-1}$.

Consideremos o sistema $A_{m \times m} \cdot X_{m \times m} + B_{m \times p} \cdot Y_{p \times m} = I_{m \times m}$, $C_{p \times m} \cdot X_{m \times m} + D_{p \times p} \cdot Y_{p \times m} = 0_{p \times m}$, X , Y e 0 são matrizes $m \times m$, $p \times m$ e $p \times m$, respectivamente. Como $A \cdot X + B \cdot Y = I$, então $X + A^{-1} \cdot B \cdot Y = A^{-1}$, ou seja, $X = A^{-1} - A^{-1} \cdot B \cdot Y$. Substituindo na segunda equação, resulta $C \cdot (A^{-1} - A^{-1} \cdot B \cdot Y) + D \cdot Y = C \cdot A^{-1} - C \cdot A^{-1} \cdot B \cdot Y + D \cdot Y = 0$, ou seja, $C \cdot A^{-1} + (-C \cdot A^{-1} \cdot B + D) \cdot Y = 0$, portanto, $Y = -(C \cdot A^{-1} \cdot B + D)^{-1} \cdot C \cdot A^{-1}$. Nesse caso, $X = A^{-1} + A^{-1} \cdot B \cdot (-C \cdot A^{-1} \cdot B + D)^{-1} \cdot C \cdot A^{-1}$.

Os dois sistemas acima representados matricialmente são: $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} X_{m \times m} \\ Y_{p \times m} \end{pmatrix} = \begin{pmatrix} I_{m \times m} \\ 0_{p \times m} \end{pmatrix}$ e $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} X_{m \times m} \\ Y_{p \times m} \end{pmatrix} = \begin{pmatrix} I_{m \times m} \\ 0_{p \times m} \end{pmatrix}$, combinando as soluções desses sistemas, concluímos que $\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} -A^{-1} \cdot B \cdot (-C \cdot A^{-1} \cdot B + D)^{-1} & A^{-1} + A^{-1} \cdot B \cdot (-C \cdot A^{-1} \cdot B + D)^{-1} \cdot C \cdot A^{-1} \\ (-C \cdot A^{-1} \cdot B + D)^{-1} & -(-C \cdot A^{-1} \cdot B + D)^{-1} \cdot C \cdot A^{-1} \end{pmatrix}$.

Consideremos a matriz simétrica $A = (a_{ij}) = \begin{pmatrix} a_{12} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$ e a matriz $B = (b_{ij}) = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix}$, $m < n$. Suponhamos que a forma quadrática $x' \cdot A \cdot x$ está restrita ao subespaço $B \cdot x = 0$.

Seja a seguinte partição da matriz A : $A = (a_{ij}) = \begin{pmatrix} a_{12} & \dots & a_{1m} & a_{1(m+1)} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} & a_{m(m+1)} & \dots & a_{mn} \\ a_{(m+1)1} & \dots & a_{(m+1)m} & a_{(m+1)(m+1)} & \dots & a_{(m+1)n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} & a_{n(m+1)} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, A_{11} é matriz $m \times m$, A_{12} é matriz $m \times (n-m)$, A_{21} é matriz $(n-m) \times m$ e A_{22} é matriz $(n-m) \times (n-m)$. A_{11} e A_{22} são matrizes simétricas e $A_{21} = A'_{12}$.

Seja a seguinte partição da matriz B : $B = (b_{ij}) = \left(\begin{array}{ccc|ccc} b_{11} & \cdots & b_{1m} & b_{1(m+1)} & \cdots & b_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{m1} & \cdots & b_{mm} & b_{m(m+1)} & \cdots & b_{mn} \end{array} \right) = (B_1 \mid B_2)$, B_1 , $m \times m$, é tal que $|B_1| \neq 0$, B_2 é matriz $m \times (n-m)$.

Seja $B \cdot x = 0$, $B \cdot x = B \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_m \\ v_{m+1} \\ \vdots \\ v_n \end{pmatrix} = (B_1 \mid B_2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = B_1 \cdot x_1 + B_2 \cdot x_2 = 0 \in \mathbb{R}^m$, $\begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} = x_1 \in \mathbb{R}^m$, $\begin{pmatrix} v_{m+1} \\ \vdots \\ v_n \end{pmatrix} = x_2 \in \mathbb{R}^{n-m}$. Logo, $B_1 \cdot x_1 = -B_2 \cdot x_2$,

ou, $x_1 = -B_1^{-1} \cdot B_2 \cdot x_2$. Como

$$x' \cdot A \cdot x = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -B_1^{-1} \cdot B_2 \cdot x_2 & x_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} -B_1^{-1} \cdot B_2 \cdot x_2 \\ x_2 \end{pmatrix}, \quad x' \cdot A \cdot x = \begin{pmatrix} -(B_1^{-1} \cdot B_2 \cdot x_2)' & x_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} -B_1^{-1} \cdot B_2 \cdot x_2 \\ x_2 \end{pmatrix},$$

$$x' \cdot A \cdot x = \begin{pmatrix} -x_2' \cdot (B_1^{-1} \cdot B_2)' & x_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} -B_1^{-1} \cdot B_2 \cdot x_2 \\ x_2 \end{pmatrix}, \quad x' \cdot A \cdot x = x_2' \cdot \begin{pmatrix} -(B_1^{-1} \cdot B_2)' & I_{n-m} \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} -B_1^{-1} \cdot B_2 \\ I_{n-m} \end{pmatrix} \cdot x_2.$$

Se $D = \begin{pmatrix} -(B_1^{-1} \cdot B_2)' & I_{n-m} \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} -B_1^{-1} \cdot B_2 \\ I_{n-m} \end{pmatrix}$, então

$D = (B_1^{-1} \cdot B_2)' \cdot A_{11} \cdot (B_1^{-1} \cdot B_2) - (B_1^{-1} \cdot B_2)' \cdot A_{12} - A_{21} \cdot (B_1^{-1} \cdot B_2) + A_{22}$. Como A_{11} e A_{22} são simétricas e $A_{21} = A_{12}'$, então $D = D'$, isto é, D é simétrica. Como $x' \cdot A \cdot x = x_2' \cdot D \cdot x_2$, então a forma quadrática $x' \cdot A \cdot x$, restrita ao subespaço $B \cdot x = 0$, é igual à forma $x_2' \cdot D \cdot x_2$ no $\mathbb{R}^{n-m}$, se

$$x = \begin{pmatrix} v_1 \\ \vdots \\ v_m \\ v_{m+1} \\ \vdots \\ v_n \end{pmatrix} \text{ e } x_2 = \begin{pmatrix} v_{m+1} \\ \vdots \\ v_n \end{pmatrix}. \text{ Portanto, a definição de } x' \cdot A \cdot x \text{ em } B \cdot x = 0 \text{ é a mesma}$$

definição de $x_2' \cdot D \cdot x_2$ no $\mathbb{R}^{n-m}$, pois para todo $x = \begin{pmatrix} v_1 \\ \vdots \\ v_m \\ v_{m+1} \\ \vdots \\ v_n \end{pmatrix}$, tal que $B \cdot x = 0$, existe um

$x_2 = \begin{pmatrix} v_{m+1} \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^{n-m}$, tal que $x' \cdot A \cdot x = x_2' \cdot D \cdot x_2$; e para todo $x_2 = \begin{pmatrix} v_{m+1} \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^{n-m}$, existe $x_1 = -B_1^{-1} \cdot B_2 \cdot x_2$, tal que $B \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = B \cdot x = 0$, $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ e $x' \cdot A \cdot x = x_2' \cdot D \cdot x_2$. Por exemplo, A é positiva-definida em $B \cdot x = 0$ se, e só se, D é positiva-definida em $\mathbb{R}^{n-m}$.

Seja $E = \begin{pmatrix} B_1 & 0 & B_2 \\ A_{11} & B_1' & A_{12} \\ A_{21} & B_2' & A_{22} \end{pmatrix}$, então $|E| = |B_1| \cdot \left| \begin{pmatrix} B_1' & A_{12} \\ B_2' & A_{22} \end{pmatrix} - \begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix} \cdot B_1^{-1} \cdot \begin{pmatrix} 0 & B_2 \end{pmatrix} \right|$, pelo complemento de Schur. Logo, $|E| = |B_1| \cdot \left| \begin{pmatrix} B_1' & A_{12} - A_{11} \cdot B_1^{-1} \cdot B_2 \\ B_2' & A_{22} - A_{21} \cdot B_1^{-1} \cdot B_2 \end{pmatrix} \right|$. $|E| = |B_1| \cdot |B_1'| \cdot |A_{22} - A_{21} \cdot B_1^{-1} \cdot B_2 - B_2' \cdot (B_1')^{-1} \cdot (A_{12} - A_{11} \cdot B_1^{-1} \cdot B_2)|$, novamente pelo complemento de Schur. Portanto, $|E| = |B_1|^2 \cdot |A_{22} - A_{21} \cdot (B_1^{-1} \cdot B_2) - (B_1^{-1} \cdot B_2)' \cdot A_{12} + (B_1^{-1} \cdot B_2)' \cdot A_{11} \cdot (B_1^{-1} \cdot B_2)| = |B_1|^2 \cdot |D|$.

Se consideramos a matriz $C = \left(\begin{array}{cc|c} A_{11} & A_{12} & B'_1 \\ A_{21} & A_{22} & B'_2 \\ B_1 & B_2 & 0 \end{array} \right)$, então $|C| = \left| \begin{array}{cc|c} A_{11} & A_{12} & B'_1 \\ A_{21} & A_{22} & B'_2 \\ B_1 & B_2 & 0 \end{array} \right| = (-1)^m \cdot \left| \begin{array}{c|cc} B_1 & 0 & B_2 \\ A_{21} & B'_2 & A_{22} \end{array} \right| = (-1)^m \cdot E$.

Portanto, $(-1)^m \cdot |C| = |B_1|^2 \cdot |D|$. Logo, $(-1)^m \cdot \left| \begin{array}{cc|c} A_{11} & A_{12} & B'_1 \\ A_{21} & A_{22} & B'_2 \\ B_1 & B_2 & 0 \end{array} \right| > 0$ se e só se $|D| > 0$.

A definição da matriz orlada (*bordered*) de A com os coeficientes de

B é $H_n = \left(\begin{array}{c|c} A & B' \\ B & 0_m \end{array} \right) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & b_{11} & \cdots & b_{m1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} & b_{1n} & \cdots & b_{mn} \\ b_{11} & \cdots & b_{1n} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} & 0 & \cdots & 0 \end{array} \right)$. As submatrizes menores principais líderes

H_r , $r = m+1, \dots, n$, que nos interessam, são:

$$H_{m+1} = \left(\begin{array}{c|c} A_{m+1} & B_{m+1}' \\ B_{m+1}' & 0_m \end{array} \right) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1(m+1)} & b_{11} & \cdots & b_{m1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{(m+1)1} & \cdots & a_{(m+1)(m+1)} & b_{1(m+1)} & \cdots & b_{m(m+1)} \\ b_{11} & \cdots & b_{1(m+1)} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{m(m+1)} & 0 & \cdots & 0 \end{array} \right), \dots, H_r = \left(\begin{array}{c|c} A_r & B_r' \\ B_r' & 0_m \end{array} \right) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1r} & b_{11} & \cdots & b_{m1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rr} & b_{1r} & \cdots & b_{mr} \\ b_{11} & \cdots & b_{1r} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mr} & 0 & \cdots & 0 \end{array} \right), \dots,$$

$r = m+1, \dots, n$. Observe que são $n-m$ determinantes H_r .

Acima mostramos que $(-1)^m \cdot |H_n| > 0$ se, e só se, $|D| > 0$, ou seja, mostramos que o hessiano orlado de ordem $n+m$ da matriz H_n e o menor principal líder de ordem $n-m$ da matriz D são tais que $(-1)^m \cdot |H_n| = |B_1|^2 \cdot |D_{n-m}|$. Estamos usando a notação $|D_{n-m}|$ para o menor líder de ordem $n-m$, ou seja, $|D|$.

Nosso objetivo é mostrar que essa relação se mantém para os demais menores líderes: $(-1)^m \cdot |H_r| = |B_1|^2 \cdot |D_{r-m}|$, $r = m+1, \dots, n$, isto é, $(-1)^m \cdot |H_{m+1}| = |B_1|^2 \cdot |D_1|, \dots, (-1)^m \cdot |H_{n-1}| = |B_1|^2 \cdot |D_{n-m-1}|$.

Vejamos para o caso $r = m+1$. Consideremos o menor líder A_{m+1} de

ordem $r = m+1$ da matriz A : $A_{m+1} = \left(\begin{array}{ccc|c} a_{12} & \cdots & a_{1m} & a_{1(m+1)} \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mm} & a_{m(m+1)} \\ \hline a_{(m+1)1} & \cdots & a_{(m+1)m} & a_{(m+1)(m+1)} \end{array} \right) = \left(\begin{array}{c|c} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right)$, A_{11} é matriz $m \times m$, A_{12} é

matriz $m \times 1$, $A_{21} = A'_{12}$ é matriz $1 \times m$. Consideremos também a matriz B :

$B = (b_{ij}) = \left(\begin{array}{ccc|c} b_{11} & \cdots & b_{1m} & b_{1(m+1)} \\ \vdots & \ddots & \vdots & \vdots \\ b_{m1} & \cdots & b_{mm} & b_{m(m+1)} \end{array} \right) = (B_1 | B_2)$, B_1 , $m \times m$, B_2 é matriz $m \times 1$. Como anteriormente,

$(-1)^m \cdot |C_1| = |B_1|^2 \cdot |D_1|$, tais que $D_1 = (B_1^{-1} \cdot B_2)' \cdot A_{11} \cdot (B_1^{-1} \cdot B_2) - (B_1^{-1} \cdot B_2)' \cdot A_{12} - A_{21} \cdot (B_1^{-1} \cdot B_2) + A_{22}$ é um número, pois

$A_{22} = a_{(m+1)(m+1)}$, $A_{21} = (a_{(m+1)1} \cdots a_{(m+1)m})$, $A_{21} = \begin{pmatrix} a_{1(m+1)} \\ \vdots \\ a_{m(m+1)} \end{pmatrix}$, $B_1^{-1} \cdot B_2 = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{pmatrix}^{-1} \cdot \begin{pmatrix} b_{1(m+1)} \\ \vdots \\ b_{m(m+1)} \end{pmatrix}$. Portanto, D_1 é o menor

líder de ordem um da matriz da D acima e $C_1 =$

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} & a_{1(m+1)} & b_{11} & \cdots & b_{m1} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} & a_{m(m+1)} & b_{1m} & \cdots & b_{mm} \\ a_{(m+1)1} & \cdots & a_{(m+1)m} & a_{(m+1)(m+1)} & b_{1(m+1)} & \cdots & b_{m(m+1)} \\ b_{11} & \cdots & b_{1m} & b_{1(m+1)} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} & b_{m(m+1)} & 0 & \cdots & 0 \end{pmatrix}.$$

Portanto, $(-1)^m \cdot |H_{m+1}| = |B_1|^2 \cdot |D_1|$. Todas as demais expressões são obtidas de maneira semelhante.

Teorema 13

A matriz simétrica $n \times n$, $|B_m| \neq 0$.

i) $x' \cdot A \cdot x > 0$, $\forall x \neq 0$, tal que $B \cdot x = 0$ e $|B_1| \neq 0$, se, e somente se, $(-1)^m \cdot \begin{vmatrix} A_r & B'_r \\ B_m & 0_m \end{vmatrix} > 0$, $r = m+1, \dots, n$;

ii) $x' \cdot A \cdot x < 0$, $\forall x \neq 0$, tal que $B \cdot x = 0$ e $|B_1| \neq 0$ se, e somente se, $(-1)^r \cdot \begin{vmatrix} A_r & B'_r \\ B_m & 0_m \end{vmatrix} > 0$, $r = m+1, \dots, n$.

Prova: i) $x' \cdot A \cdot x > 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e só se, $x' \cdot D \cdot x > 0$, $\forall x \neq 0, x \in \mathbb{R}^{n-m}$. Como $(-1)^m \cdot |H_r| = |B_1|^2 \cdot |D_{r-m}|$, $r = m+1, \dots, n$, então $(-1)^m \cdot |H_r| > 0$ se, e só se, $|D_{r-m}| > 0$, $r = m+1, \dots, n$, o que implica que $x' \cdot A \cdot x > 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e só se, $(-1)^m \cdot |H_r| > 0$, $r = m+1, \dots, n$.

ii) $x' \cdot A \cdot x < 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e só se, $x' \cdot D \cdot x < 0$, $\forall x \neq 0, x \in \mathbb{R}^{n-m}$. $x' \cdot D \cdot x < 0$, $\forall x \neq 0, x \in \mathbb{R}^{n-m}$ se, e só se, $(-1)^k \cdot |D_{m+k-m}| > 0$, $k = 1, \dots, n-m$, isto, é $(-1)^{r-m} \cdot |D_{r-m}| > 0$, $r = m+1, \dots, n$. Portanto, $x' \cdot A \cdot x < 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e só se, $(-1)^r \cdot |H_r| > 0$, $r = m+1, \dots, n$.

Teorema 14

x^* é ponto crítico do lagrangeano $\mathcal{L}(x) = f(x) - \lambda_1 \cdot (g_1(x) - c_1) - \dots - \lambda_m \cdot (g_m(x) - c_m)$ do problema:

$$\begin{aligned} \min/\max \quad & f(x_1, x_2, \dots, x_n) \\ \text{sa} \quad & g_1(x_1, x_2, \dots, x_n) = c_1 \\ & \vdots \\ & g_m(x_1, x_2, \dots, x_n) = c_m. \end{aligned}$$

$$\text{Seja } |H_r(x^*)| = \begin{vmatrix} \mathcal{L}''(x^*)_{r \times r} & \nabla g_1(x^*)_{r \times 1} & \cdots & \nabla g_m(x^*)_{r \times 1} \\ \nabla g_1(x^*)'_{1 \times r} & & & \\ \vdots & & & \\ \nabla g_m(x^*)'_{1 \times r} & & 0_{m \times m} & \end{vmatrix} =$$

$$\left| \begin{array}{ccc|ccc} \frac{\partial^2 \mathcal{L}''}{\partial x_1^2}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial x_r \partial x_1}(x^*) & \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_m}{\partial x_1}(x^*) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}''}{\partial x_1 \partial x_r}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial x_r^2}(x^*) & \frac{\partial g_1}{\partial x_r}(x^*) & \dots & \frac{\partial g_m}{\partial x_r}(x^*) \\ \hline \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_1}{\partial x_r}(x^*) & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(x^*) & \dots & \frac{\partial g_m}{\partial x_r}(x^*) & 0 & \dots & 0 \end{array} \right|, r = m+1, \dots, n.$$

i) Se $(-1)^m \cdot |H_r(x^*)| > 0$, então x^* é um mínimo local para o problema acima;

ii) Se $(-1)^r \cdot |H_r(x^*)| > 0$, então x^* é um máximo local para o problema acima.

Primeiramente, observe que caso o número m de restrições seja par, então $|H_{m+1}(x^*)| > 0, \dots, |H_n(x^*)| > 0$ implica que x^* é um mínimo local; caso o número m de restrições seja ímpar, então $|H_{m+1}(x^*)| < 0, \dots, |H_n(x^*)| < 0$ implica que x^* é um mínimo local; caso o número m de restrições seja par, então $|H_{m+1}(x^*)| < 0, |H_{m+2}(x^*)| > 0, \dots, (-1)^n \cdot |H_n(x^*)| > 0$ implica que x^* é um máximo local; caso o número m de restrições seja ímpar, então $|H_{m+1}(x^*)| > 0, |H_{m+2}(x^*)| < 0, \dots, (-1)^n \cdot |H_n(x^*)| > 0$ implica que x^* é um máximo local.

A segunda observação é que

$$|H_r(x^*)| = \frac{\left| \frac{\partial^2 \mathcal{L}''}{\partial x_j \partial x_i}(x^*) \right| \left| -\frac{\partial^2 \mathcal{L}''}{\partial \lambda_j \partial x_i}(x^*) \right|}{\left| \frac{\partial^2 \mathcal{L}''}{\partial x_j \partial \lambda_i}(x^*) \right| \left| 0 \right|} = \frac{\left| \frac{\partial^2 \mathcal{L}''}{\partial x_j \partial x_i}(x^*) \right| \left| \frac{\partial^2 \mathcal{L}''}{\partial \lambda_j \partial x_i}(x^*) \right|}{\left| \frac{\partial^2 \mathcal{L}''}{\partial x_j \partial \lambda_i}(x^*) \right| \left| 0 \right|} =$$

$$\left| \begin{array}{ccc|ccc} \frac{\partial^2 \mathcal{L}''}{\partial x_1^2}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial x_r \partial x_1}(x^*) & \frac{\partial^2 \mathcal{L}''}{\partial \lambda_1 \partial x_1}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial \lambda_m \partial x_1}(x^*) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}''}{\partial x_1 \partial x_r}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial x_r^2}(x^*) & \frac{\partial^2 \mathcal{L}''}{\partial \lambda_1 \partial x_r}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial \lambda_m \partial x_r}(x^*) \\ \hline \frac{\partial^2 \mathcal{L}''}{\partial x_1 \partial \lambda_1}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial x_r \partial \lambda_1}(x^*) & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}''}{\partial x_1 \partial \lambda_m}(x^*) & \dots & \frac{\partial^2 \mathcal{L}''}{\partial x_r \partial \lambda_m}(x^*) & 0 & \dots & 0 \end{array} \right|$$

x_i, x_j , tais que $i, j = 1, \dots, r$, λ_i, λ_j , tais que $i, j = 1, \dots, m$.

Exemplo 21: Ao dizermos que x^* é um mínimo ou máximo local na restrição, por exemplo, $g(x) = c$, a proposição deve ser interpretada de maneira rigorosa. Seja o problema:

$$\begin{aligned} \min \quad & x^2 + 2 \cdot y^2 + x \cdot y + 1 \\ \text{sa} \quad & x + y = 1. \end{aligned}$$

O lagrangeano desse problema é $\mathcal{L} = x^2 + 2 \cdot y^2 + x \cdot y + 1 - \mu \cdot (x + y - 1)$. O ponto crítico é $\left(\frac{3}{4}, \frac{1}{4}\right)$ com $\mu = \frac{7}{4}$. Como $(-1)^1 \cdot |H_2| = \begin{vmatrix} 2 & 2 & 2 \\ 1 & 4 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 4 > 0$, então $\left(\frac{3}{4}, \frac{1}{4}\right)$ é um **mínimo local** em $x + y - 1 = 0$.

Como $f(x, y) = x^2 + 2 \cdot y^2 + x \cdot y + 1$ é uma função estritamente convexa, então, pelo Teorema de Kuhn-Tucker, $\left(\frac{3}{4}, \frac{1}{4}\right)$ é um **mínimo global estrito** em $x + y - 1 \geq 0$ e, por consequência, mínimo global estrito na restrição $x + y - 1 = 0$.

Entretanto, se entrarmos para o interior da região $x + y - 1 \leq 0$ através da trajetória $\varphi(t) = \left(\frac{3}{4} - t, \frac{1}{4} - t\right)$, obtemos $f(\varphi(t)) = 4 \cdot t^2 - \frac{7}{2} \cdot t + \frac{15}{8} < f\left(\frac{3}{4}, \frac{1}{4}\right)$, pois $f(\varphi(t))$ decresce para $0 < t \leq \frac{7}{16}$. Então, não existe uma vizinhança em torno de $\left(\frac{3}{4}, \frac{1}{4}\right)$ na qual esse ponto seja um mínimo.

Exemplo 22: Seja

$$\begin{array}{ll} \min/\max & x \cdot y - x^3 - y^3 - \frac{1}{27} \\ \text{sa} & x - y = 0. \end{array}$$

$\mathcal{L} = x \cdot y - x^3 - y^3 - \frac{1}{27} - \lambda \cdot (x - y)$, então $\frac{\partial \mathcal{L}}{\partial x} = y - 3 \cdot x^2 - \lambda = 0$, $\frac{\partial \mathcal{L}}{\partial y} = x - 3 \cdot y^2 + \lambda = 0$. Como $x = y$, as soluções dessas equações são $(0, 0)$, $\lambda = 0$ e $(1/3, 1/3)$, $\lambda = 0$. $|H_2(x, y)| = \begin{vmatrix} -6 \cdot x & 1 & 1 \\ 1 & -6 \cdot y & -1 \\ 1 & -1 & 0 \end{vmatrix} = 6 \cdot x + 6 \cdot y - 2$, então $(-1)^1 \cdot |H_2(0, 0)| = 2$ e $(-1)^2 \cdot |H_2(1/3, 1/3)| = 2$. Portanto, $(0, 0)$ é um mínimo local e $(1/3, 1/3)$ é um máximo local para o problema acima.

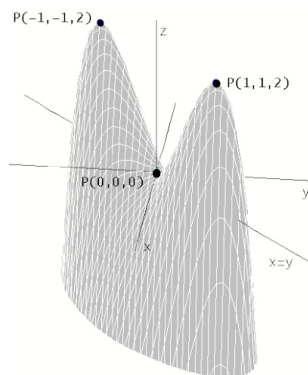
Exemplo 23: Seja

$$\begin{array}{ll} \min/\max & 4 \cdot x \cdot y - x^4 - y^4 \\ \text{sa} & x - y = 0. \end{array}$$

$\mathcal{L} = 4 \cdot x \cdot y - x^4 - y^4 - \lambda(x - y)$, então $\frac{\partial \mathcal{L}}{\partial x} = 4 \cdot y - 4 \cdot x^3 - \lambda = 0$, $\frac{\partial \mathcal{L}}{\partial y} = x - 4 \cdot y^3 + \lambda = 0$. Como $x = y$, as soluções dessas equações são $(-1, -1)$, $\lambda = 0$, $(0, 0)$, $\lambda = 0$, e $(1, 1)$, $\lambda = 0$.

$|H_2(x, y)| = \begin{vmatrix} -12 \cdot x^2 & 4 & 1 \\ 4 & -12 \cdot y^2 & -1 \\ 1 & -1 & 0 \end{vmatrix} = 12 \cdot x^2 + 12 \cdot y^2 - 8$ $|H_2(-1, -1)| = |H_2(1, 1)| = 16 > 0$, portanto, $(-1, -1)$ e $(1, 1)$ são máximos locais; $(-1) \cdot |H_2(0, 0)| = 8$, então $(0, 0)$ é um mínimo local. Se examinarmos a função $f(x, y) = 4 \cdot x \cdot y - x^4 - y^4$, concluiremos que $(-1, -1)$ e $(1, 1)$ são

máximos globais de f e que $(0,0)$ não é nem máximo e nem mínimo. Veja o desenho. Pela definição dada, $(0,0)$ é uma sela, pois $f''(0,0) = \begin{vmatrix} 0 & 4 \\ 4 & 0 \end{vmatrix}$:



Entretanto, sobre a reta $x=y$, $(0,0)$ é um mínimo global.

Exemplo 24: Seja a função $f(x, y, z) = x^3 - 2 \cdot x \cdot y - y^2 + z^2$. Os pontos críticos de f são $(0,0,0)$ e $(-2/3, 2/3, 0)$. A derivada segunda de f é $f'' = \begin{pmatrix} 6 \cdot x & -2 & 0 \\ -2 & -2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. A derivada

$f''(0,0,0) = \begin{pmatrix} 0 & -2 & 0 \\ -2 & -2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ é indefinida, o que implica que $(0,0,0)$ é uma sela. Digamos que restringimos f à reta determinada pelos planos $x-y=0$ e $z=0$, então $f = x^3 - 3 \cdot x^2 = x^2 \cdot (x-3)$, facilmente observamos que $(0,0,0)$ é um máximo local nessa direção. Veja o desenho abaixo. Observe que a restrição é

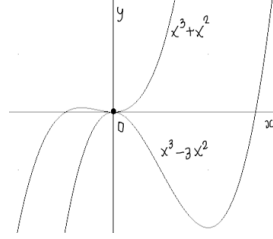
$B \cdot v = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0$. Então, $H = \begin{pmatrix} 6 \cdot x & -2 & 0 & 1 & 0 \\ -2 & -2 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$. Como $(-1)^3 \cdot \begin{vmatrix} 0 & -2 & 0 & 1 & 0 \\ -2 & -2 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{vmatrix} = 6 > 0$, $f''(0,0,0)$ é

negativa-definida na direção da reta $x=y$ e $z=0$, portanto, $(0,0,0)$ é um máximo local nessa direção.

Se restringirmos f à reta $y=-x$ e $z=0$, então $f = x^3 + x^2 = x^2 + 1$ e $(0,0,0)$ é um

mínimo local. Como $(-1)^2 \cdot \begin{vmatrix} 0 & -2 & 0 & 1 & 0 \\ -2 & -2 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{vmatrix} = 2 > 0$, $f''(0,0,0)$ é positiva-definida na

direção da reta $y=-x$ e $z=0$, portanto, $(0,0,0)$ é um mínimo local nessa direção.



$f''(-2/3, 2/3, 0) = \begin{pmatrix} -4 & -2 & 0 \\ -2 & -2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ é indefinida, então $(-2/3, 2/3, 0)$ é uma sela. Mas, claramente vemos que sobre a reta $x+y=0, z=0$, ou seja, sobre $(s, -s, 0), s \in \mathbb{R}$, $f(-s, s, 0) = s^3 + s^2$. O ponto $s = -\frac{2}{3}$ é um máximo local para $s^3 + s^2$, isso implica que $(-2/3, 2/3, 0)$ é um máximo local para f quando restrita à reta $x+y=0, z=0$.

Observe que $(-1)^3 \cdot \begin{vmatrix} -4 & -2 & 0 & 1 & 0 \\ -2 & -2 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{vmatrix} = 2 > 0$.

Exemplo 25: Seja

$$\begin{array}{ll} \min/\max & 4 \cdot x^2 \cdot y + 3 \cdot x \cdot y^2 - 12 \cdot x \cdot y \\ \text{sa} & x - y = 0. \end{array}$$

$\mathcal{L} = 4 \cdot x^2 \cdot y + 3 \cdot x \cdot y^2 - 12 \cdot x \cdot y - \lambda \cdot (x - y)$, então $\frac{\partial \mathcal{L}}{\partial x} = 8 \cdot x \cdot y + 3 \cdot y^2 - 12 \cdot y - \lambda = 0$, $\frac{\partial \mathcal{L}}{\partial y} = 4 \cdot x^2 + 6 \cdot x \cdot y - 12 \cdot x + \lambda = 0$. $(0, 0)$, $\lambda = 0$, é um ponto crítico do lagrangeano e $|H_2(x, y)| = \begin{vmatrix} 8 \cdot y & 8 \cdot x + 6 \cdot y - 12 & 1 \\ 8 \cdot x + 6 \cdot y - 12 & 6 \cdot x & -1 \\ 1 & -1 & 0 \end{vmatrix} = -22 \cdot x - 20 \cdot y + 24$, então $(-1)^2 \cdot |H_2(0, 0)| = 24$, portanto, na direção da reta $y = x$, $(0, 0)$ é um máximo local.

Considere o problema:

$$\begin{array}{ll} \min/\max & 4 \cdot x^2 \cdot y + 3 \cdot x \cdot y^2 - 12 \cdot x \cdot y \\ \text{sa} & x + y = 0. \end{array}$$

$\mathcal{L} = 4 \cdot x^2 \cdot y + 3 \cdot x \cdot y^2 - 12 \cdot x \cdot y - \lambda \cdot (x + y)$, então $\frac{\partial \mathcal{L}}{\partial x} = 8 \cdot x \cdot y + 3 \cdot y^2 - 12 \cdot y - \lambda = 0$, $\frac{\partial \mathcal{L}}{\partial y} = 4 \cdot x^2 + 6 \cdot x \cdot y - 12 \cdot x - \lambda = 0$. $(0, 0)$, $\lambda = 0$, é um ponto crítico do lagrangeano e $|H_2(x, y)| = \begin{vmatrix} 8 \cdot y & 8 \cdot x + 6 \cdot y - 12 & 1 \\ 8 \cdot x + 6 \cdot y - 12 & 6 \cdot x & 1 \\ 1 & 1 & 0 \end{vmatrix} = 10 \cdot x + 4 \cdot y - 24$, então $(-1)^1 \cdot |H_2(0, 0)| = 24 > 0$, portanto, na direção da reta $y = -x$, $(0, 0)$ é um mínimo local. Observe que $f''(x, y) = \begin{vmatrix} 8 \cdot y & 8 \cdot x + 6 \cdot y - 12 \\ 8 \cdot x + 6 \cdot y - 12 & 6 \cdot x \end{vmatrix}$; $f''(0, 0) = \begin{vmatrix} 0 & -12 \\ -12 & 0 \end{vmatrix}$ é indefinida.

Exercício 6

Em relação à função $f(x, y) = 4 \cdot x^2 \cdot y + 3 \cdot x \cdot y^2 - 12 \cdot x \cdot y$, observe que $(0, 4)$, $(3, 0)$ e $(1, 4/3)$ são pontos críticos; $(0, 4)$ e $(3, 0)$ são selas; e $(1, 4/3)$ é um mínimo local. Mostre

que na sela $(0, 4)$, $H_2 = \begin{vmatrix} 8 \cdot y & 8 \cdot x + 6 \cdot y - 12 & 2 \\ 8 \cdot x + 6 \cdot y - 12 & 6 \cdot x & 1 \\ 2 & 1 & 0 \end{vmatrix} > 0$; conclua que sobre a reta $2 \cdot x + y = 4$,

$(0, 4)$ é máximo local. Mostre que na sela $(0, 4)$, $H_2 = \begin{vmatrix} 8 \cdot y & 8 \cdot x + 6 \cdot y - 12 & 1 \\ 8 \cdot x + 6 \cdot y - 12 & 6 \cdot x & -2 \\ 1 & -2 & 0 \end{vmatrix} < 0$; conclua que sobre a reta $x - 2 \cdot y = -8$, $(0, 4)$ é mínimo local.

O resultado anterior também se aplica para os menores principais das matrizes $\begin{vmatrix} A & B' \\ B & 0 \end{vmatrix}$ e D .

Exemplo 26: Consideremos $H = \begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \\ b_1 & b_2 & b_3 & 0 \end{pmatrix}$, então $A_{11} = (a_{11})$, $A_{12} = (a_{12} \ a_{13})$, $A_{21} = \begin{pmatrix} a_{21} \\ a_{31} \end{pmatrix}$, $A_{22} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}$, $B_1 = (b_1)$ e $B_2 = (b_2 \ b_3)$. Logo, $D = \frac{a_{11}}{b_1^2} \cdot \begin{pmatrix} b_2 \\ b_3 \end{pmatrix} \cdot (b_2 \ b_3) - \frac{1}{b_1} \cdot \begin{pmatrix} a_{21} \\ a_{31} \end{pmatrix} \cdot (b_2 \ b_3) - \frac{1}{b_1} \cdot \begin{pmatrix} b_2 \\ b_3 \end{pmatrix} \cdot (a_{12} \ a_{13}) + \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}$ ou

$$D = \begin{pmatrix} \frac{a_{11} \cdot b_2^2}{b_1^2} - \frac{(a_{21} + a_{12}) \cdot b_2}{b_1} + a_{22} & \frac{a_{11} \cdot b_2 \cdot b_3}{b_1^2} - \frac{a_{21} \cdot b_3 + b_2 \cdot a_{13}}{b_1} + a_{23} \\ \frac{a_{11} \cdot b_2 \cdot b_3}{b_1^2} - \frac{a_{31} \cdot b_2 + b_3 \cdot a_{12}}{b_1} + a_{32} & \frac{a_{11} \cdot b_3^2}{b_1^2} - \frac{(a_{31} + a_{13}) \cdot b_3}{b_1} + a_{33} \end{pmatrix}.$$

Como já vimos, $-|H_3| = -\begin{vmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \\ b_1 & b_2 & b_3 & 0 \end{vmatrix} = |b_1|^2 \cdot |D_2| = |b_1|^2 \cdot |D|$. Vejamos os menores de

ordem um de D : $|b_1|^2 \cdot |D_1| = a_{11} \cdot b_2^2 - a_{12} \cdot b_2 \cdot b_1 - a_{21} \cdot b_2 \cdot b_1 + a_{22} \cdot b_1^2 = -\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ b_1 & b_2 & 0 \end{vmatrix} = -|H_2|$;

$|b_1|^2 \cdot |D_{22}| = a_{11} \cdot b_3^2 - a_{13} \cdot b_3 \cdot b_1 - a_{31} \cdot b_3 \cdot b_1 + a_{33} \cdot b_1^2 = -\begin{vmatrix} a_{11} & a_{13} & b_1 \\ a_{31} & a_{33} & b_3 \\ b_1 & b_3 & 0 \end{vmatrix}$. Portanto, o menor de ordem um líder da matriz D corresponde ao menor que chamamos de $|H_2|$, e o outro menor de ordem um da matriz D corresponde ao menor em que a linha dois e a coluna dois da matriz H foram eliminadas.

Consideremos a situação geral $\begin{vmatrix} A_{11} & \cdots & A_{1m} & B'_1 \\ \vdots & \ddots & \vdots & \vdots \\ A_{m1} & \cdots & A_{mm} & B'_m \\ B_1 & \cdots & B_m & 0 \end{vmatrix} = \begin{vmatrix} A_{11} & \cdots & A_{1m} & B'_1 \\ \vdots & A_{22}(i_1, \dots, i_k) & \vdots & \vdots \\ B_1 & \cdots & B_m & 0 \end{vmatrix}$, $\begin{vmatrix} a_{11} & \cdots & a_{1m} & a_{1h} & \cdots & a_{1h_k} & b_{11} & \cdots & b_{m1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} & a_{mh} & \cdots & a_{mh_k} & b_{1m} & \cdots & b_{mm} \\ a_{h1} & \cdots & a_{hm} & a_{hh} & \cdots & a_{hh_k} & b_{1h} & \cdots & b_{mh} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{km} & a_{kh} & \cdots & a_{kh_k} & b_{1k} & \cdots & b_{mk} \\ b_{11} & \cdots & b_{1m} & b_{1h} & \cdots & b_{1h_k} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} & b_{mh} & \cdots & b_{mh_k} & 0 & \cdots & 0 \end{vmatrix},$

que consiste em eliminar colunas das matrizes A_{12} e B_2 , conservando as colunas $i_1, \dots, i_k$, eliminar linhas das matrizes A_{21} e B'_2 , conservando as linhas $i_1, \dots, i_k$. Observe que as matrizes A_{11} , B_1 e B'_1 são conservadas. Observe

também que os menores de interesse são menores da matriz determinados pelos menores da matriz A_{22} . Por exemplo, $|H_2| = |(A_{22})_1|, \dots, |H_r| = |(A_{22})_r|$. A matriz A_{22} , cuja ordem é $n-m$, tem $2^{n-m}-1$ menores. O mesmo número de menores acima, que podemos notar por $|H(i_1, \dots, i_k)|$, ao número de menores da matriz D , cuja ordem também é $n-m$.

No exemplo acima, o menor $-|H(3,3)| = -\begin{vmatrix} a_{11} & a_{13} & b_{11} \\ a_{31} & a_{33} & b_{13} \\ b_{11} & b_{13} & 0 \end{vmatrix} = b_{11}^2 \cdot |D(2,2)|$. Então, observe que $(-1)^m \cdot |H(i_1, \dots, i_k)| = |B_1|^2 \cdot |D(i_1-m, \dots, i_k-m)|$.

Como a definição da forma $x' \cdot A \cdot x$, tal que $B \cdot x = 0$, e a mesma definição de $x' \cdot D \cdot x, x \in \mathbb{R}^{n-m}$, então

Teorema 15

i) Seja A simétrica $n \times n$, $|B_m| \neq 0$. $x' \cdot A \cdot x \geq 0, \forall x$, tal que $B' \cdot x = 0$ e $|B_1| \neq 0$, se, e somente se, $(-1)^m \cdot |H(i_1, \dots, i_k)| \geq 0$ para todos os menores principais;

ii) Seja A simétrica $n \times n$, $|B_m| \neq 0$. $x' \cdot A \cdot x \leq 0, \forall x$, tal que $B' \cdot x = 0$, se, e somente se, $|B_1| \neq 0$, se, e somente se, $(-1)^k \cdot |H(i_1, \dots, i_k)| \geq 0$ para todos os menores principais.

Uma observação importante em relação ao teorema acima é que a notação comumente empregada, devida a Debreu, não é nada clara! Por um rearranjo principal, digamos π , que não modifique as posições iniciais $1, \dots, m$ e traga para as posições $m+1, \dots, m+r$ as $i_1, \dots, i_r$ linhas e colunas da matriz A , que faz com que $A(i_1, \dots, i_k)$ ocupe a posição líder na matriz A_{22} e, portanto, ao considerarmos o determinante resultante de eliminar as demais linhas e colunas de A_{22} é $|H(i_1, \dots, i_k)|$. Portanto, para obter todos os menores $|H(i_1, \dots, i_k)|$ devemos aplicar todos os rearranjos principais π desde que não altere as posições iniciais $m+1, \dots, m+k$.

Os determinantes $\begin{vmatrix} A_r^\pi & B_{rm}^\pi \\ B_{rm}^{\pi'} & 0_m \end{vmatrix}, r = m+1, \dots, n$, são os menores obtidos da matriz A^π , em que $A_r^\pi = \begin{pmatrix} a_{11} & \dots & a_{1m} & a_{1i_1} & \dots & a_{1i_r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} & a_{mi_1} & \dots & a_{mi_r} \\ a_{i_1 1} & \dots & a_{i_1 m} & a_{i_1 i_1} & \dots & a_{i_1 i_r} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i_r 1} & \dots & a_{i_r m} & a_{i_r i_1} & \dots & a_{i_r i_r} \end{pmatrix}$ e B_{rm}^π é a matriz $B_{mr}^\pi = \begin{pmatrix} b_{11} & \dots & b_{1m} & b_{1i_1} & \dots & b_{1i_r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mm} & b_{mi_1} & \dots & b_{mi_r} \end{pmatrix}$, tal que as colunas conservadas são as mesmas que foram conservadas na

matriz A^x . De fato, $\begin{vmatrix} A_r^x & B_{rm}^x \\ B_{rm}^{x'} & 0_m \end{vmatrix}$ é uma notação muito pouco esclarecedora. É conveniente observar atentamente o que são as matrizes A_r^x , B_{mr}^x e

$$\begin{vmatrix} A_r^x & B_{rm}^x \\ B_{rm}^{x'} & 0_m \end{vmatrix} = \left(\begin{array}{cccccc|cccc} a_{i_1} & \cdots & a_{1m} & a_{1i_1} & \cdots & a_{1i_r} & b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} & a_{mi_1} & \cdots & a_{mi_r} & b_{m1} & \cdots & b_{mm} \\ a_{i_1} & \cdots & a_{i_1m} & a_{i_1i_1} & \cdots & a_{i_1i_r} & b_{i_11} & \cdots & b_{i_1m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{i_r1} & \cdots & a_{i_rm} & a_{i_ri_1} & \cdots & a_{i_ri_r} & b_{i_r1} & \cdots & b_{i_rm} \\ \hline b_{11} & \cdots & b_{m1} & b_{i_11} & \cdots & b_{i_r1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{1m} & \cdots & b_{mm} & b_{i_1m} & \cdots & b_{i_rm} & 0 & \cdots & 0 \end{array} \right).$$

Nessa notação, o Teorema de Mann-Debreu se enuncia como:

Teorema 16 (Critério de Mann-Debreu)

A simétrica $n \times n$, $|B_m| = \begin{vmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{vmatrix} \neq 0$, $m < n$.

- i) $x' \cdot A \cdot x > 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e somente se $(-1)^m \cdot |H_r| = (-1)^m \cdot \begin{vmatrix} A_r^x & B_{rm}^x \\ B_{rm}^{x'} & 0_m \end{vmatrix} > 0$, $r = m+1, \dots, n$;
- ii) $x' \cdot A \cdot x < 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e somente se $(-1)^r \cdot |H_r| = (-1)^r \cdot \begin{vmatrix} A_r^x & B_{rm}^x \\ B_{rm}^{x'} & 0_m \end{vmatrix} > 0$, $r = m+1, \dots, n$;
- iii) $x' \cdot A \cdot x \geq 0$, $\forall x$ tal que $B' \cdot x = 0$, se, e somente se, $(-1)^m \cdot |H_r^x| = (-1)^m \cdot \begin{vmatrix} A_r^x & B_{rm}^x \\ B_{rm}^{x'} & 0_m \end{vmatrix} \geq 0$, $r = m+1, \dots, n$ e $\forall \pi$.

- iv) $x' \cdot A \cdot x \leq 0$, $\forall x$ tal que $B' \cdot x = 0$, se, e somente se, $(-1)^r \cdot |H_r^x| = (-1)^r \cdot \begin{vmatrix} A_r^x & B_{rm}^x \\ B_{rm}^{x'} & 0_m \end{vmatrix} \geq 0$, $r = m+1, \dots, n$ e $\forall \pi$.

Na situação geral em que o menor de ordem m de B diferente de zero não é o $B_1 = \begin{pmatrix} b_{11} & & b_{1m} \\ & \ddots & \\ b_{m1} & & b_{mm} \end{pmatrix}$, mas, digamos, $|B_m| = |B(j_1, \dots, j_m)| = \begin{vmatrix} b_{1j_1} & & b_{1j_m} \\ & \ddots & \\ b_{mj_1} & & b_{mj_m} \end{vmatrix} \neq 0$, então podemos usar um rearranjo principal π que traga as posições $j_1, \dots, j_m$ para as posições iniciais $1, \dots, m$. Nesse caso, a matriz A_{22}^x são as demais linhas e colunas da matriz A e são os menores A_{22}^x orlados pelas correspondentes linhas e colunas da matriz B que interessam.

Exemplo 27: Digamos que $b_{12} \neq 0$, então $|H| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & b_{11} \\ a_{21} & a_{22} & a_{23} & b_{12} \\ a_{31} & a_{32} & a_{33} & b_{13} \\ b_{11} & b_{12} & b_{13} & 0 \end{vmatrix} = \begin{vmatrix} a_{22} & a_{21} & a_{23} & b_{12} \\ a_{12} & a_{11} & a_{23} & b_{11} \\ a_{32} & a_{31} & a_{33} & b_{13} \\ b_{12} & b_{11} & b_{13} & 0 \end{vmatrix}$. Os

menores que devemos avaliar são $\begin{vmatrix} a_{22} & a_{21} & b_{12} \\ a_{12} & a_{11} & b_{11} \\ b_{12} & b_{11} & 0 \end{vmatrix}$, $\begin{vmatrix} a_{22} & a_{21} & b_{12} \\ a_{12} & a_{33} & b_{13} \\ b_{12} & b_{13} & 0 \end{vmatrix}$ e $\begin{vmatrix} a_{22} & a_{21} & a_{23} & b_{12} \\ a_{12} & a_{11} & a_{23} & b_{11} \\ a_{32} & a_{31} & a_{33} & b_{13} \\ b_{12} & b_{11} & b_{13} & 0 \end{vmatrix}$.

Exemplo 28: Voltemos ao problema da determinação do ponto sobre o parabolóide $g(x, y, z) = x^2 + y^2 - z = 0$, que está mais próximo do ponto $P(0, 0, 1)$:

$$\begin{array}{ll} \min & x^2 + y^2 + (z-1)^2 \\ \text{sa} & x^2 + y^2 - z = 0. \end{array}$$

A derivada segunda do lagrangeano $\mathcal{L}(x, y, z) = x^2 + y^2 + (z-1)^2 - \lambda \cdot (x^2 + y^2 - z)$ é $\mathcal{L}''(x, y, z) = \begin{pmatrix} 2 \cdot (1-\lambda) & 0 & 0 \\ 0 & 2 \cdot (1-\lambda) & 0 \\ 0 & 0 & 2 \end{pmatrix}$ e $\nabla g(x, y, z) = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ -1 \end{pmatrix}$. Os pontos de mínimo estão sobre a

circunferência $x^2 + y^2 = \frac{1}{2}, z = \frac{1}{2}$, $\lambda^* = 1$. Como $|H_3| = \begin{vmatrix} 2 \cdot (1-\lambda) & 0 & 0 & 2 \cdot x \\ 0 & 2 \cdot (1-\lambda) & 0 & 2 \cdot y \\ 0 & 0 & 2 & -1 \\ 2 \cdot x & 2 \cdot y & -1 & 0 \end{vmatrix}$, então nos pontos

sobre a circunferência acima $|H_3| = \begin{vmatrix} 0 & 0 & 0 & 2 \cdot x \\ 0 & 0 & 0 & 2 \cdot y \\ 0 & 0 & 2 & -1 \\ 2 \cdot x & 2 \cdot y & -1 & 0 \end{vmatrix} = 0$. Por exemplo,

consideremos os pontos $\left(\pm \frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$, então $|H_3| = \begin{vmatrix} 0 & 0 & 0 & \pm\sqrt{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ \pm\sqrt{2} & 0 & -1 & 0 \end{vmatrix} = 0$, $\begin{vmatrix} 0 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 0 \end{vmatrix} = 0$ e

$|H_3| = \begin{vmatrix} 0 & 0 & \pm\sqrt{2} \\ 0 & 2 & -1 \\ \pm\sqrt{2} & -1 & 0 \end{vmatrix} = (-1)^1 \cdot 4$. Portanto, todos os menores $(-1)^m \cdot |H_r''| \geq 0$, isto é, a forma

quadrática $\mathcal{L}''\left(\pm \frac{\sqrt{2}}{2}, 0, \frac{1}{2}\right)$ é positiva-semidefinida nos planos ortogonais aos vetores $v = \begin{pmatrix} \pm\sqrt{2} \\ 0 \\ -1 \end{pmatrix}$.

Se consideremos todos os pontos sobre a circunferência, então

$|H_3| = \begin{vmatrix} 0 & 0 & 0 & 2 \cdot x \\ 0 & 0 & 0 & 2 \cdot y \\ 0 & 0 & 2 & -1 \\ 2 \cdot x & 2 \cdot y & -1 & 0 \end{vmatrix} = 0$, $\begin{vmatrix} 0 & 0 & 2 \cdot y \\ 0 & 2 & -1 \\ 2 \cdot y & -1 & 0 \end{vmatrix} = (-1)^1 \cdot (-8 \cdot y^2) \geq 0$ e $|H_3| = \begin{vmatrix} 0 & 0 & 2 \cdot x \\ 0 & 2 & -1 \\ 2 \cdot x & -1 & 0 \end{vmatrix} = (-1)^1 \cdot (-8 \cdot x^2) \geq 0$. Portanto, todos

os menores $(-1)^m \cdot |H_r''| \geq 0$, isto é, a forma quadrática $\mathcal{L}''\left(x, y, \frac{1}{2}\right)$ é positiva-semidefinida nos planos ortogonais aos vetores $v = \begin{pmatrix} 2 \cdot x \\ 2 \cdot y \\ -1 \end{pmatrix}$.

Exemplo 29: Analisemos o problema $f(x, y, z) = x + y - x \cdot y - 2 \cdot y \cdot z$, sujeita às restrições $g(x, y, z) = 2 \cdot x + y + z = 4$ e $h(x, y, z) = x + 2 \cdot y + z^2 = 4$. O ponto $(1, 1, 1)$ satisfaz as duas restrições e é um ponto crítico do lagrangeano

$$\mathcal{L} = x + y - x \cdot y - 2 \cdot y \cdot z - \lambda_1 \cdot (2 \cdot x + y + z - 4) - \lambda_2 \cdot (x + 2 \cdot y + z^2 - 4);$$

$$\frac{\partial \mathcal{L}}{\partial x} = 1 - y - 2 \cdot \lambda_1 = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = 1 - x - 2 \cdot z - \lambda_1 - 2 \cdot \lambda_2 = 0,$$

$$\frac{\partial \mathcal{L}}{\partial z} = -2 \cdot y - \lambda_1 - 2 \cdot \lambda_2 \cdot z = 0. \text{ Observe que } \lambda_1 = \frac{2}{3} \text{ e } \lambda_2 = -\frac{4}{3}. \text{ O ponto } (1, 1, 1) \text{ é uma sela da}$$

função f , pois $f'' = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & -2 \\ 0 & -2 & 0 \end{pmatrix}$ é uma forma indefinida. Na curva definida pela interseção das superfícies $g(x, y, z) = 2 \cdot x + y + z = 4$ e $h(x, y, z) = x + 2 \cdot y + z^2 = 4$, como

$$|H_3| = \begin{vmatrix} 0 & -1 & 0 & 2 & 1 \\ -1 & 0 & -2 & 1 & 2 \\ 0 & -2 & -2\mu & 1 & 2 \cdot z \\ 2 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 \cdot z & 0 & 0 \end{vmatrix}, \text{ então } (-1)^2 \cdot |H_3(1, 1, 1)| = \begin{vmatrix} 0 & -1 & 0 & 2 & 1 \\ -1 & 0 & -2 & 1 & 2 \\ 0 & -2 & \frac{8}{3} & 1 & 2 \\ 2 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 0 \end{vmatrix} = (-1)^2 \cdot 60 > 0. \text{ Logo, } (1, 1, 1) \text{ é um ponto}$$

de mínimo local de f quando restrita à curva $g(x, y, z) = 2 \cdot x + y + z = 4$ e $h(x, y, z) = x + 2 \cdot y + z^2 = 4$.

Exercício 7

Seja a função $f(x, y, z) = \frac{x^2}{2} + \frac{y^2}{2} + \frac{z^2}{2}$ sujeita às restrições $2 \cdot x + y + z = 2$ e $x + y + z = 6$. Mostre que $(-4, 5, 5)$ é um mínimo local. Na verdade, é um mínimo global. Por quê?

Exercício 8

Seja a função $f(x, y, z) = -x^2 - y^2 - z^2 + x \cdot y + z$ sujeita à restrição $x + y + z = 35$. Mostre que $(\frac{69}{5}, \frac{69}{5}, \frac{37}{5})$ é um máximo local, calculando $|H_2| = \begin{vmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & 0 \end{vmatrix}$ e $|H_3| = \begin{vmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & 0 & 1 \\ 0 & 0 & -2 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix}$. Na verdade, $(\frac{69}{5}, \frac{69}{5}, \frac{37}{5})$ é um máximo global. Por quê?

Exercício 9

Seja a função $f(x, y, z) = 4 \cdot z - x^2 - y^2 - z^2$ sujeita às restrições $z - x \cdot y = 0$ e $x^2 + y^2 + z^2 = 3$. Se $\mathcal{L} = 4 \cdot z - x^2 - y^2 - z^2 - \lambda_1 \cdot (z - x \cdot y) - \lambda_2 \cdot (x^2 + y^2 + z^2 - 3)$, então os pontos críticos do lagrangeano e respectivos multiplicadores são $(1, 1, 1)$, $\lambda_1 = 2$, $\lambda_2 = 0$; $(1, -1, -1)$, $\lambda_1 = 2$, $\lambda_2 = -2$; $(-1, 1, -1)$, $\lambda_1 = 2$,

$$\lambda_2 = -2 \text{ e } (-1, -1, 1), \quad \lambda_1 = 2, \quad \lambda_2 = 0. \text{ Considerando } |H_3| = \begin{vmatrix} -2 \cdot (\mu + 1) & \lambda & 0 & -y & 2 \cdot x \\ \lambda & -2 \cdot (\mu + 1) & 0 & -x & 2 \cdot y \\ 0 & 0 & -2 \cdot (\mu + 1) & 1 & 2 \cdot z \\ -y & -x & 1 & 0 & 0 \\ 2 \cdot x & 2 \cdot y & 2 \cdot z & 0 & 0 \end{vmatrix},$$

determine entre os pontos acima quais são mínimos e quais são máximos locais de f sujeita às restrições.

Exercício 10

Seja a função $f(x, y, z) = x^2 + y^2 + z^2$ sujeita às restrições $x - y + z = 1$ e $2 \cdot x + y - z = 3$. Se $\mathcal{L} = x^2 + y^2 + z^2 - \lambda_1 \cdot (x - y + z - 1) - \lambda_2 \cdot (2 \cdot x + y - z - 3)$, então o ponto crítico do lagrangeano e respectivos multiplicadores são $\left(\frac{4}{3}, \frac{1}{6}, -\frac{1}{6}\right)$, $\lambda_1 = \frac{2}{3}$, $\lambda_2 = 1$, $\lambda_3 = 0$. Considerando

$$|H_3| = \begin{vmatrix} 2 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & -1 & 1 \\ 0 & 0 & 2 & 1 & -1 \\ 1 & -1 & 1 & 0 & 0 \\ 2 & 1 & -1 & 0 & 0 \end{vmatrix}, \text{ mostre que } \left(\frac{4}{3}, \frac{1}{6}, -\frac{1}{6}\right) \text{ é um mínimo local. Na verdade, } \left(\frac{4}{3}, \frac{1}{6}, -\frac{1}{6}\right)$$

é um mínimo global, pois é o ponto sobre a reta determinada por $x - y + z = 1$ e $2 \cdot x + y - z = 3$ que está mais próximo da origem.

Exercício 11

Seja a função $f(x, y, z) = x + 2 \cdot y + z$ sujeita à restrição $x^2 + y^2 - z = 0$. Se $\mathcal{L} = x + 2 \cdot y + z - \lambda \cdot (x^2 + y^2 - z)$, mostre que $\left(-\frac{1}{2}, -1, \frac{5}{4}\right)$, $\lambda = -1$, é um mínimo local.

7. Teorema de Arrow-Enthoven

O Teorema 5 de Arrow e Enthoven¹⁰ estabelece um critério suficiente para a quase-concavidade de uma função definida em $\mathbb{R}_+^n$. O critério utiliza os chamados hessianos orlados.

Mudemos a notação para as derivadas parciais, de maneira que não fique excessivamente complexa: a derivada $\frac{\partial f}{\partial x_i}$ notamos por f_i e a derivada segunda $\frac{\partial^2 f}{\partial x_j \partial x_i}$, por f_{ij} . A derivada $\frac{\partial f}{\partial x_i}(x^*)$ notamos por f_i^* . Seja $f: \mathbb{R}_+^n \rightarrow \mathbb{R}$, definimos os hessianos orlados:

$$D_1 = \begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix}, \quad D_2 = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix}, \dots, \quad D_r = \begin{vmatrix} 0 & f_1 & \dots & f_r \\ f_1 & f_{11} & \dots & f_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ f_r & f_{r1} & \dots & f_{rr} \end{vmatrix} < 0, \quad \dots, \quad D_n = \begin{vmatrix} 0 & f_1 & \dots & f_n \\ f_1 & f_{11} & \dots & f_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ f_n & f_{n1} & \dots & f_{nn} \end{vmatrix}. \quad \text{É comum}$$

apresentar esses determinantes como $D_r = \begin{vmatrix} f_{11} & \dots & f_{1r} & f_1 \\ \vdots & \ddots & \vdots & \vdots \\ f_{r1} & \dots & f_{rr} & f_r \\ f_1 & \dots & f_r & 0 \end{vmatrix} < 0, \quad r = 1, \dots, n$.

Exemplo 30: Seja $f: \mathbb{R}_+^3 \rightarrow \mathbb{R}$. Suponhamos que $D_1 = \begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix} = -(f_1)^2 < 0$ e $D_2 = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix} > 0$.

Consideremos o problema:

$$\begin{aligned} \max \quad & f(x) \\ \text{sa} \quad & \nabla f(x^*) \cdot (x - x^*) \leq 0 \\ & x \geq 0, \end{aligned}$$

ou seja, o problema de determinar o máximo de f na região do $\mathbb{R}^3$ determinada por $\nabla f(x^*) \cdot (x - x^*) \leq 0$ e $x \geq 0$.

Primeiramente, mostramos que x^* é um máximo local estrito para f no plano $\nabla f(x^*) \cdot (x - x^*) = 0$, ortogonal ao $\nabla f(x^*)$ que contém x^* . Não vamos usar o lagrangeano para resolver esse problema. Vamos expressar o problema através de uma função $F: \mathbb{R}_+^2 \rightarrow \mathbb{R}$ de duas variáveis, definida em $\mathbb{R}_+^2$, e mostramos que x^* é um máximo local para F e, por consequência, um máximo local para f numa certa vizinhança de x^* restrita ao plano $\nabla f(x^*) \cdot (x - x^*) = 0$.

Como, por hipótese, $\frac{\partial f}{\partial x_1}(x^*) = f_1^* \neq 0$, então para os pontos $x = (x_1, x_2, x_3)$ sobre $\nabla f(x^*) \cdot (x - x^*) = 0$, podemos expressar x_1 em função das variáveis x_2 e x_3 :

$x_1(x_2, x_3) = \frac{1}{\left(\frac{\partial f}{\partial x_1}(x^*)\right)} \cdot \left(\nabla f(x^*) \cdot x^* - \left(\frac{\partial f}{\partial x_2}(x^*) \cdot x_2 + \frac{\partial f}{\partial x_3}(x^*) \cdot x_3 \right) \right)$. Não existe nada complicado nisto: $\frac{\partial f}{\partial x_1}(x^*) \cdot x_1 + \frac{\partial f}{\partial x_2}(x^*) \cdot x_2 + \frac{\partial f}{\partial x_3}(x^*) \cdot x_3 = \nabla f(x^*) \cdot x^*$ é um plano definido no $\mathbb{R}^3$ e como $\frac{\partial f}{\partial x_1}(x^*) \neq 0$, podemos expressar x_1 em função de x_2 e x_3 , por exemplo, em $2 \cdot x_1 + 5 \cdot x_2 - 7 \cdot x_3 = -2$; podemos expressar $x_1 = -\frac{5}{2} \cdot x_2 + \frac{7}{2} \cdot x_3 - 1$, em que $(x_2, x_3) \in \mathbb{R}^2$. No caso, estamos restringindo o domínio ao $\mathbb{R}_+^2$. Portanto, substituindo x_1 em $f(x_1, x_2, x_3)$, resulta na função $F(x_2, x_3) = f(x_1(x_2, x_3), x_2, x_3)$ e $F: \mathbb{R}_+^2 \rightarrow \mathbb{R}$.

As derivadas de x_1 são $\frac{\partial x_1}{\partial x_i}(x_2, x_3) = -\frac{\frac{\partial f}{\partial x_i}(x^*)}{\frac{\partial f}{\partial x_1}(x^*)}$, $i = 2, 3$. No exemplo, $\frac{\partial x_1}{\partial x_2} = -\frac{5}{2}$ e $\frac{\partial x_1}{\partial x_3} = \frac{7}{2}$. As derivadas parciais de F , para x_i , $i = 2, 3$, são:

$$\begin{aligned} \frac{\partial F}{\partial x_i}(x_2, x_3) &= \\ \frac{\partial f}{\partial x_1}(x_1, x_2, x_3) \cdot \frac{\partial x_1}{\partial x_i} + \frac{\partial f}{\partial x_i}(x_1, x_2, x_3) &= \frac{\partial f}{\partial x_1}(x_1, x_2, x_3) \cdot \left(-\frac{\frac{\partial f}{\partial x_i}(x^*)}{\frac{\partial f}{\partial x_1}(x^*)} \right) + \frac{\partial f}{\partial x_i}(x_1, x_2, x_3). \end{aligned}$$

Na nova notação, por exemplo, $F_2 = f_1 \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_2$ e $F_3 = f_1 \cdot \left(-\frac{f_3^*}{f_1^*} \right) + f_3$.

Se derivarmos mais uma vez, obtemos

$$\frac{\partial^2 F}{\partial x_2^2} = \frac{\partial^2 f}{\partial x_1^2} \cdot \left(-\frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial f}{\partial x_1}(x^*)} \right)^2 + \frac{\partial^2 f}{\partial x_2 \partial x_1} \cdot \left(-\frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial f}{\partial x_1}(x^*)} \right) + \frac{\partial^2 f}{\partial x_1 \partial x_2} \cdot \left(-\frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial f}{\partial x_1}(x^*)} \right) + \frac{\partial^2 f}{\partial x_2^2},$$

que na nova notação é

$$F_{22} = f_{11} \cdot \left(-\frac{f_2^*}{f_1^*} \right)^2 + f_{12} \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{21} \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{22}.$$

As outras derivadas são

$$F_{23} = f_{11} \cdot \left(-\frac{f_3^*}{f_1^*} \right) \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{13} \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{21} \cdot \left(-\frac{f_3^*}{f_1^*} \right) + f_{23},$$

$$F_{32} = f_{11} \cdot \left(-\frac{f_2^*}{f_1^*} \right) \cdot \left(-\frac{f_3^*}{f_1^*} \right) + f_{12} \cdot \left(-\frac{f_3^*}{f_1^*} \right) + f_{31} \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{32}$$

e

$$F_{33} = f_{11} \cdot \left(-\frac{f_3^*}{f_1^*} \right)^2 + f_{13} \cdot \left(-\frac{f_3^*}{f_1^*} \right) + f_{31} \cdot \left(-\frac{f_3^*}{f_1^*} \right) + f_{33}.$$

Seja o determinante $c_2 = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix} > 0$. Se dividirmos a primeira coluna

por f_1 , obtemos $f_1 \cdot \begin{vmatrix} 0 & f_1 & f_2 \\ 1 & f_{11} & f_{12} \\ \frac{f_2}{f_1} & f_{21} & f_{22} \end{vmatrix}$. A seguir multiplicamos a segunda linha do

determinante por $-\frac{f_2}{f_1}$ e somamos na terceira linha. Essa operação não

altera o determinante inicial e resulta em $f_1 \cdot \begin{vmatrix} 0 & f_1 & f_2 \\ 1 & f_{11} & f_{12} \\ 0 & f_{21} + \left(-\frac{f_2}{f_1} \right) \cdot f_{11} & f_{22} + \left(-\frac{f_2}{f_1} \right) \cdot f_{12} \end{vmatrix}$.

Desenvolvendo o determinante pela primeira coluna, obtemos

$$(-f_1) \cdot \begin{vmatrix} f_1 & f_2 \\ f_{21} + \left(-\frac{f_2}{f_1} \right) \cdot f_{11} & f_{22} + \left(-\frac{f_2}{f_1} \right) \cdot f_{12} \end{vmatrix}.$$

Dividimos a primeira linha por f_1 , resultando em

$$-(f_1)^2 \cdot \begin{vmatrix} 1 & \frac{f_2}{f_1} \\ f_{21} + \left(-\frac{f_2}{f_1} \right) \cdot f_{11} & f_{22} + \left(-\frac{f_2}{f_1} \right) \cdot f_{12} \end{vmatrix}.$$

Multiplicamos a primeira coluna por $-\frac{f_2}{f_1}$ e

somamos na segunda coluna. O determinante não se modifica:

$$\begin{aligned} & -(f_1)^2 \cdot \begin{vmatrix} 1 & 0 \\ f_{21} + \left(-\frac{f_2}{f_1} \right) \cdot f_{11} & f_{11} \cdot \left(-\frac{f_2}{f_1} \right)^2 + f_{21} \cdot \left(-\frac{f_2}{f_1} \right) + f_{22} + \left(-\frac{f_2}{f_1} \right) \cdot f_{12} \end{vmatrix} = \\ & -(f_1)^2 \cdot \left(f_{11} \cdot \left(-\frac{f_2}{f_1} \right)^2 + f_{12} \cdot \left(-\frac{f_2}{f_1} \right) + f_{21} \cdot \left(-\frac{f_2}{f_1} \right) + f_{22} \right). \end{aligned}$$

Esses cálculos dependem apenas do fato que $f_1(x) \neq 0$, para todo $x \in \mathbb{R}_+^3$, e das propriedades dos determinantes.

Se nos restringirmos ao ponto x^* , então

$$C_2^* = \begin{vmatrix} 0 & f_1^* & f_2^* \\ f_1^* & f_{11}^* & f_{12}^* \\ f_2^* & f_{21}^* & f_{22}^* \end{vmatrix} = -(f_1^*)^2 \cdot \left(f_{11}^* \cdot \left(-\frac{f_2^*}{f_1^*} \right)^2 + f_{12}^* \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{21}^* \cdot \left(-\frac{f_2^*}{f_1^*} \right) + f_{22}^* \right) = -(f_1^*)^2 \cdot F_{22}^*. \text{ Portanto, } F_{22}^* < 0.$$

Considere o determinante $C_3^* = \begin{vmatrix} 0 & f_1^* & f_2^* & f_3^* \\ f_1^* & f_{11}^* & f_{12}^* & f_{13}^* \\ f_2^* & f_{21}^* & f_{22}^* & f_{23}^* \\ f_3^* & f_{31}^* & f_{32}^* & f_{33}^* \end{vmatrix} < 0$. Se dividirmos a primeira

coluna por f_1^* , obtemos $f_1^* \cdot \begin{vmatrix} 0 & f_1^* & f_2^* & f_3^* \\ 1 & f_{11}^* & f_{12}^* & f_{13}^* \\ \frac{f_2^*}{f_1^*} & f_{21}^* & f_{22}^* & f_{23}^* \\ \frac{f_3^*}{f_1^*} & f_{31}^* & f_{32}^* & f_{33}^* \end{vmatrix}$. A seguir multiplicamos a segunda

linha do determinante por $-\frac{f_2^*}{f_1^*}$, somamos na terceira linha e depois multiplicamos a segunda linha por $-\frac{f_3^*}{f_1^*}$ e somamos na quarta linha. Essas operações não alteram o determinante inicial, resultando em Desenvolvendo o determinante pela primeira coluna, obtemos

. A seguir vamos dividir a primeira linha por f_1^* :

. Multiplicamos a primeira coluna por $-\frac{f_2^*}{f_1^*}$ e somamos na segunda coluna; multiplicamos a primeira coluna por $-\frac{f_3^*}{f_1^*}$ e somamos na terceira coluna. Essas operações não alteram o determinante e resultam em

.

Portanto, . Como $F_{22}^* < 0$ e , então F é localmente estritamente côncava.

Observe que $e_{F_3} = f_1^* \cdot \left(-\frac{f_3^*}{f_1^*}\right) + f_3^*$, então e . Logo, e é um ponto crítico de F e, portanto, um máximo único para F localmente, isto é, para na proximidade de e e tal que $\nabla f(x^*) \cdot (x - x^*) = 0$.

Mostremos que x^* é um máximo global para f na região em determinada por $\nabla f(x^*) \cdot (x - x^*) \leq 0$. Seja , tal que . Definimos , para . Se e é o segmento de reta unindo x^* e , então . tem, necessariamente, um ponto de mínimo .

Como , se , então , pois implicaria que e é um máximo local em . Se , então . Num caso ou outro, se o mínimo está em , então . Como , então .

Logo, . Mas, vejamos, na primeira parte mostramos que x^* é um máximo local para f quando restrita a uma vizinhança do plano

$\nabla f(x^*) \cdot (x - x^*) = 0$ e x^* é um ponto qualquer em $\mathbb{R}^3$ escolhido arbitrariamente, então, igualmente, x^* é um máximo local estrito no plano π . Portanto, o mínimo de f não pode estar em π . Sendo assim, o mínimo está em $\partial\pi$, ou seja, $\partial\pi$.

Logo, provamos que para qualquer x^* , se $\nabla f(x^*) \cdot (x - x^*) \leq 0, \forall x \in \pi$, então x^* é um máximo local estrito de f em π . Isso é suficiente para assegurar que f é quase-côncava. Vejamos, sejam dois pontos $x_1, x_2 \in \pi$ e uma combinação convexa desses dois pontos $\lambda x_1 + (1-\lambda)x_2$, isto é, $\lambda x_1 + (1-\lambda)x_2 \in \pi$, então $\nabla f(x^*) \cdot (\lambda x_1 + (1-\lambda)x_2 - x^*) \leq 0$, portanto, $\nabla f(x^*) \cdot (\lambda(x_1 - x^*) + (1-\lambda)(x_2 - x^*)) \leq 0$. Então, $\lambda \nabla f(x^*) \cdot (x_1 - x^*) + (1-\lambda) \nabla f(x^*) \cdot (x_2 - x^*) \leq 0$. Desde que $\nabla f(x^*) \cdot (x_1 - x^*) \leq 0$ e $\nabla f(x^*) \cdot (x_2 - x^*) \leq 0$, então se $\lambda \geq 0$ e $1-\lambda \geq 0$, então $\lambda \nabla f(x^*) \cdot (x_1 - x^*) + (1-\lambda) \nabla f(x^*) \cdot (x_2 - x^*) \leq 0$. Em qualquer situação, $\nabla f(x^*) \cdot (\lambda x_1 + (1-\lambda)x_2 - x^*) \leq 0$, o que implica que f é quase-côncava.

Por outro lado, se f é quase-côncava, então as condições de Kuhn-Tucker, se satisfeitas, são suficientes para assegurar que x^* é um máximo para f em relação ao problema

O lagrangeano do problema acima é A 1ª condição de K-T,

é satisfeita em x^* com $\lambda^* = 1$, então x^* é máximo global para f em π e, portanto, para a $F(x_2, x_3) = f(x_1(x_2, x_3), x_2, x_3)$ definida anteriormente, então F é positiva-semidefinida. Logo,

e

.

Observe que x^* é um ponto qualquer, então $x^* \in \mathbb{R}^3$. Se ocorrer f , a condição acima é automaticamente satisfeita.

O resultado anterior pode ser empregado em relação às funções quase-convexas. Facilmente, constatamos que f é quase-convexa se e só se f é quase-côncava, pois, em relação a qualquer função, os contornos inferiores de f são os contornos superiores da função $-f$. Além disto, obviamente, f é quase-convexa se e só se $-f$ é quase-côncava.

Teorema 17

Seja f .

- i) Suponhamos que f é quase-côncava em $\mathbb{R}_+^1$;
- ii) Se f é quase-côncava em $\mathbb{R}_+^1$, então $-f$ é quase-convexa em $\mathbb{R}_+^1$;
- iii) Suponhamos que $-f$ é quase-convexa em $\mathbb{R}_+^1$;
- iv) Se $-f$ é quase-convexa em $\mathbb{R}_+^1$, então f é quase-côncava em $\mathbb{R}_+^1$.

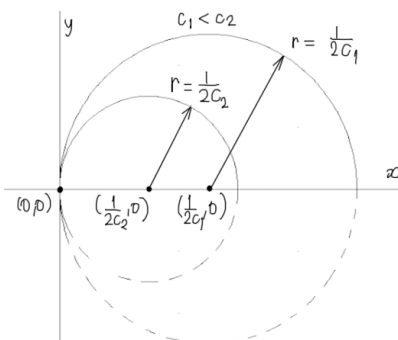
O resultado acima pode ser estendido para $\mathbb{R}_+^n$ para obter o

Teorema 18 – Teorema de Arrow-Enthoven

Seja f .

- i) Suponhamos que $D_1 = \begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix} = -(f_1)^2 < 0$, $D_2 = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix} > 0$, ..., então f é quase-côncava em $\mathbb{R}_+^n$;
- ii) Se f é quase-côncava em $\mathbb{R}_+^n$, então $-f$ é quase-convexa em $\mathbb{R}_+^n$;
- iii) Suponhamos que $D_1 = \begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix} = -(f_1)^2 < 0$, ..., então $-f$ é quase-convexa em $\mathbb{R}_+^n$;
- iv) Se $-f$ é quase-convexa em $\mathbb{R}_+^n$, então f é quase-côncava em $\mathbb{R}_+^n$.

Exemplo 31: Seja $f(x, y) = x^2 + y^2$. Não é difícil ver que f é quase-côncava, pois para $x, y \geq 0$, se, e só se, $x = y = 0$, ou seja, as isoquantas são circunferências cujos centros estão nos pontos $(-x, -y)$ e os raios são $\sqrt{x^2 + y^2}$. Logo, os contornos superiores são definidos pelos círculos interceptados por $f = c$, ou seja, a parte do círculo tal que $x, y \geq 0$.



O teorema acima, nesse exemplo, é árduo de ser aplicado, pois

Observe que essa função não é nem côncava e nem convexa, pois se, e só se, .

Exemplo 32: Seja , . Como e , então f é quase-côncava. Observe que essa função não é nem côncava e nem convexa.

Exemplo 33: Seja , . Como e , então f é quase-convexa.

Observe que essa função não é nem côncava e nem convexa.

Exemplo 34: Seja , . Como e , então f é quase-convexa. Observe que essa função não é nem côncava e nem convexa. Observe que os contornos inferiores são definidos pela interseção dos círculos com , o que nos permite observar facilmente que f é quase-convexa.

Exemplo 35: Seja , . Como , e , então f é quase-côncava. Considere o problema:

então, o ponto de máximo está no ponto . No desenho abaixo, estão desenhadas: a restrição orçamentária , e a isoquanta máxima que o consumidor alcança, que é na cesta :

Exercício 12

Exercício 12.1: Seja u , é quase-concava e não é nem côncava e nem convexa.

Exercício 12.2: Considere o problema:

Mostre que u é quase-côncava e obtenha as demandas marshallianas x_1^M e x_2^M .

Exercício 12.3: Seja u , v , w . Mostre que u e v e que, portanto, $u + v$ é quase-côncava.

Mostre que u, v é quase-convexa.

Exercício 12.4: Seja u , v . Resolva o problema:

Obtenha as demandas marshallianas e a utilidade indireta $e(p, y)$.

Respostas: $x_1^M = \frac{y}{2p_1}$, $x_2^M = \frac{y}{2p_2}$. Se $y = 1$, então $x_1^M = \frac{1}{2p_1}$, $x_2^M = \frac{1}{2p_2}$.

Observe que u, v é côncava sem ser estrita, pois $u = v = 0$.

Exercício 12.5: Seja u . Mostre que u é quase-côncava e observe que u é quase-convexa. Observe que, dada a isoquanta u , então u é côncava. Portanto, u é côncava. Qual é a cesta ótima se $y = 1$. Resposta: $x_1^M = \frac{1}{2p_1}$, $x_2^M = \frac{1}{2p_2}$.

8. Critérios para Mínimos e Máximos Locais com Restrições, Desigualdades e Igualdades

Finalmente, chegamos ao problema mais geral que podemos formular:

Exemplo 36: Considere o problema de localizar o ponto que está mais próximo da origem:

Observe o desenho à direita na figura abaixo. A região é definida por e . Obviamente, a solução está no ponto :

O lagrangeano é , então, em , temos e . A fronteira é sempre ativa. Como , qualquer candidato a mínimo deve satisfazer a condição de Kuhn-Tucker para mínimo, ou seja, . $f(x, y, z) = x^2 + y^2 + z^2$ é estritamente convexa, então é um mínimo global.

Considere o problema:

Em relação à região , definida por e , o ponto é ponto crítico do lagrangeano com e e satisfaz as condições de Kuhn-Tucker para um mínimo. é um mínimo global em , pois $f(x, y, z) = x^2 + y^2 + z^2$ é estritamente convexa. Mas, o ponto , tal que e também está satisfazendo as condições de Kuhn-Tucker para um mínimo. Mas, observe que não é mínimo local nem máximo local.

Exemplo 37: Considere o problema:

Sejam , e , o ponto é um ponto crítico do lagrangeano . Os multiplicadores são e $\mu_2 = -4$. Observe que para , tal que $\nabla g(0, -1, 2) \cdot v = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \cdot v = 0$ e

$\nabla h(0, -1, 2) \cdot v = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \cdot v = 0$, ou seja, na direção do eixo x , $v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot t, t \neq 0$, pois. Portanto, ao longo da fronteira $(t, -1, 2)$, $(0, -1, 2)$ é um mínimo local estrito. De fato, .

Como $(0, -1, 2)$ satisfaz as condições de Kuhn-Tucker para mínimo, então . Seja uma trajetória $\varphi(t) = (0, -1, 2)$, para $t \geq 0$, tal que, isto é, $\varphi(t)$ é uma trajetória qualquer que entra para o interior da região determinada pela restrições, então, pois $\mu_1 > 0$ e . Portanto, , então, para valores de t e próximos de zero, $\frac{f(\varphi(t)) - f(0, -1, 2)}{t} > 0$, ou seja, . Desse modo, $(0, -1, 2)$ é um mínimo local para f quando restrita às $y \geq -1$ e $z \geq 1$. Observe que é imprescindível que $\mu_1 > 0$. Como no primeiro exemplo dessa seção, no próximo, exemplificaremos que não podemos prescindir dessa condição.

Exemplo 38: Considere o problema:

$$\begin{aligned} \min \quad & x^2 - y^2 + z^2 \\ \text{sa} \quad & -\sqrt{2} \cdot x + y + 0 \cdot z = -1 \\ & 0 \cdot x + y + z \geq 1. \end{aligned}$$

Sejam x, y, z e $h(x, y, z) = y + z \geq 1$. O ponto $(\sqrt{2}, 1, 0)$ é um ponto crítico do lagrangeano $\mathcal{L} = x^2 - y^2 + z^2 - \mu_1(-\sqrt{2} \cdot x + y + 1) - \mu_2 \cdot (y + z - 1)$. Os multiplicadores são $\mu_1 = 0$ e $\mu_2 = 0$. Observe que

para $0 \neq v \in \mathbb{R}^n$, tal que $v = \begin{pmatrix} -\frac{\sqrt{2}}{2} \cdot t \\ -1 \\ 1 \end{pmatrix} \cdot t, t \neq 0$, ou seja, na direção do eixo x , $v = \begin{pmatrix} -\frac{\sqrt{2}}{2} \cdot t \\ -1 \\ 1 \end{pmatrix} \cdot t, t \neq 0$, pois .

Portanto, ao longo da fronteira $(\sqrt{2} - \frac{\sqrt{2}}{2} \cdot t, 1 - t, t)$, $(\sqrt{2}, 1, 0)$ é um mínimo local estrito. De fato, .

Como $(\sqrt{2}, 1, 0)$ satisfaz as condições de Kuhn-Tucker para mínimo, então . Seja uma trajetória $\varphi(t) = (\sqrt{2}, 1, 0)$, $g(\varphi(t)) = -1$, para $t \geq 0$, tal que $\frac{d}{dt} h(\varphi(t)) \Big|_{t=0} = \nabla h(0, -1, 2) \cdot \varphi'(0) > 0$, isto é, $\varphi(t)$ é uma trajetória qualquer que entra para o interior da região determinada pela restrições. Entretanto, nesse caso, não podemos empregar o argumento do exemplo anterior, porque . De fato, $(\sqrt{2}, 1, 0)$ não é um mínimo local. Por exemplo, $(\sqrt{2}, 1, 0)$ é uma trajetória viável e, ou seja, nessa direção, $(\sqrt{2}, 1, 0)$ é um máximo.

Teorema 19

Consideremos o problema:

Seja $\mathcal{L} = f - \sum_{i=1}^p \lambda_i \cdot (g_i - a_i) - \sum_{j=1}^q \mu_j \cdot (h_j - b_j) - \sum_{k=1}^r \pi_k \cdot (l_k - c_k)$ o lagrangeano do problema acima e um ponto crítico desse lagrangeano.

i) Suponhamos que $v' \cdot \mathcal{L}''(x^*) \cdot v > 0$ para todo v , $\nabla h_j(x^*) \cdot v = 0$ e $\mu_j = 0$. Se x^* satisfaz a segunda e a terceira condições de Kuhn-Tucker para um mínimo, em relação às restrições $h_j(x) \geq b_j, 1 \leq j \leq q$, e tal que se as fronteiras $h_j(x) = b_j$ sejam ativas, então x^* é um mínimo local;

ii) Suponhamos que $v' \cdot \mathcal{L}''(x^*) \cdot v < 0$ para todo v , $\nabla h_j(x^*) \cdot v = 0$ e $\mu_j = 0$. Se x^* satisfaz a segunda e a terceira condições de Kuhn-Tucker para um máximo, em relação às restrições $h_j(x) \geq b_j, 1 \leq j \leq q$, e tal que se as fronteiras $h_j(x) = b_j$ sejam ativas, então x^* é um máximo local.

O resultado é evidente a partir dos exemplos e considerações acima.

Exemplo 39: Considere o problema:

Esse problema não é difícil de resolver, pois é possível expressá-lo apenas através de uma variável, por exemplo, $f(x, y, z) = x \cdot y \cdot z = -z^2 \cdot (1-z)$. Então, no intervalo $[0, 1]$, temos dois máximos globais em $z=0$ e $z=1$ e um mínimo global em $z=\frac{3}{2}$. Se $z=1$ corresponde ao ponto $P(0,0,1)$ e $f(0,0,1)=0$. Como podemos expressar $f(x, y, z) = x \cdot y \cdot z = -x^2 \cdot (1-x)$ ou $f(x, y, z) = x \cdot y \cdot z = -y^2 \cdot (1-y)$, existem outros dois máximos que são $P(1,0,0)$ e $P(0,1,0)$ e correspondem a $z=0$. O ponto $P(\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$, $z=\frac{3}{2}$, é um ponto de mínimo e, de maneira similar, são mínimos $P(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3})$ e $P(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3})$.

A interseção da esfera $x^2 + y^2 + z^2 = 1$ e do plano $x + y + z = 1$ é a circunferência centrada no ponto $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ de raio $r = \frac{\sqrt{6}}{3}$. Veja o desenho abaixo:

O lagrangeano desse problema é $\mathcal{L} = x \cdot y \cdot z - \lambda_1 \cdot (x^2 + y^2 + z^2 - 1) - \lambda_2 \cdot (x + y + z - 1)$. Usando as condições de primeira ordem do lagrangeano, obtemos os multiplicadores. Para os máximos $P(1,0,0)$, e $P(0,0,1)$ temos $\lambda_1 = 0$ e $\lambda_2 = 0$. Para os mínimos $P(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$, $P(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3})$ e $P(\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$ temos $\lambda_1 = \frac{2}{3}$ e $\lambda_2 = \frac{2}{3}$.

$$\text{Como } \mathcal{L}'' = \begin{pmatrix} -2 \cdot \lambda_1 & z & y & -2 \cdot x & -1 \\ z & -2 \cdot \lambda_1 & x & -2 \cdot y & -1 \\ y & x & -2 \cdot \lambda_1 & -2 \cdot z & -1 \\ -2 \cdot x & -2 \cdot y & -2 \cdot z & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 \end{pmatrix},$$

resulta $(-1)^3 \cdot |H_3(1,0,0)| = (-1)^3 \cdot |H_3(0,1,0)| = (-1)^3 \cdot |H_3(0,0,1)| = 8$ e, o que assegura os resultados obtidos anteriormente em relação aos máximos e mínimos locais nas restrições $x^2 + y^2 + z^2 = 1$ e $x + y + z = 1$. De fato, são máximos e mínimos globais nessas restrições.

Para o problema:

$$\begin{array}{ll} \min/\max & x \cdot y \cdot z \\ \text{sa} & x^2 + y^2 + z^2 \leq 1 \\ & x + y + z = 1, \end{array}$$

A restrição determinada por $x + y + z = 1$ é o círculo centrado no ponto $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, de raio $r = \frac{\sqrt{6}}{3}$, que está hachurado acima. Por resultados anteriores, como $\lambda_1 = \frac{2}{3}$ e $\lambda_2 = \frac{2}{3}$, então $P(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$, $P(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3})$ e $P(\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$ são mínimos locais e, como não existem outros, são globais no círculo. Em relação aos $P(1,0,0)$, e $P(0,0,1)$ não temos como afirmar nada a partir de nossos resultados. De fato, esses pontos não são de máximos ou mínimos locais em relação às restrições $x^2 + y^2 + z^2 = 1$ e $x + y + z = 1$.

O ponto $(1,0,0)$ satisfaz as condições de primeira ordem do lagrangeano com $\lambda_1 = 0$, pois a fronteira $x^2 + y^2 + z^2 = 1$ é inativa e $\lambda_2 = \frac{1}{9}$. Os resultados anteriores também não asseguram que $(1,0,0)$ é um máximo, mas, efetivamente, é o máximo global da função no plano $x + y + z = 1$ e, portanto, é máximo global em relação às restrições.

Exercício 13

Mostre que $P(4, 4, 0)$ são mínimos locais para a função $f(x, y, z) = x + 2 \cdot y + 2 \cdot z - x \cdot y - y \cdot z$, empregando

,

em que $\mathcal{L} = x + 2 \cdot y + 2 \cdot z - x \cdot y - y \cdot z - \lambda_1 \cdot (x + y + 2 \cdot z - 8) - \lambda_2 \cdot (2 \cdot x + y + z^2 - 12)$.

9. Notas históricas

9.1 Teorema de Hancock

O ponto de partida no estabelecimento de critérios para determinar a definição de formas quadráticas restritas a subespaços é o livro *“Theory of Maxima and Minima”*¹¹, de 1917, do matemático norte-americano Harris Hancock (1867-1944).

Hancock estudou na John Hopkins, depois foi para Cambridge (UK) e Berlim, onde doutorou-se. Em Berlim, Hancock estudou com Karl Weierstrass (1815-1897), o que foi determinante para ele. Os primeiros livros de Hancock após sua tese de Berlim foram: *“Lectures on the Theory of Maxima and Minima of Functions of Several Variables (Weierstrass’s Theory)”*, de 1903, e *“Lectures on the Calculus of Variations (The Weierstrassian Theory)”*, de 1904, que resultaram do estudo do trabalho de Weierstrass e do contato direto com matemáticos do círculo do grande professor alemão. Em 1892, Hancock foi contratado como professor de Cálculo das Variações em Chicago.

No começo do século 20, Chicago foi um importante centro matemático sob influência do matemático alemão Felix Klein (1849-1925). O matemático Oskar Bolza (1857-1942), que foi orientado por Klein, emigrou para os EUA e, em 1892, passou a lecionar em Chicago, onde deu origem à escola americana de Cálculo das

Variações. Hancock pertenceu a essa escola e seu trabalho foi desenvolvido a partir dos temas da escola de Weierstrass¹².

No prefácio do livro *“Theory of Maxima and Minima”*, Hancock esclarece que a teoria da otimização estática era esboçada nos cursos introdutórios de Cálculo das Variações que lecionava e que “este esboço é fundamentado nas palestras que foram apresentadas pelo falecido Professor Weierstrass na Universidade de Berlim”¹³.

Mann¹⁴ se refere a dois livros que trazem uma demonstração do Teorema de Hancock. Um deles é o *“Theory of Maxima and Minima”*, de Hancock, e outro é um livro que Mann intitula *“Determinantenrechnung”* (“Cálculo dos Determinantes”), do matemático alemão Gerhard Kowalewski (1876-1950). Não conhecemos outras referências a esse livro. De qualquer maneira, Mann se refere ao mesmo resultado apresentado por Hancock. Com alguma segurança, podemos concluir que em 1917, o ano da publicação do livro do Hancock, o que ele expôs sobre as condições para a determinação da definição de formas era o que sabia até então, seja pelo esforço de outros ou dele mesmo. Portanto, podemos concluir que esse era o estado da arte no começo do século 20.

Hancock considera o problema de minimização e maximização de uma forma quadrática $x' \cdot A \cdot x$ no sujeita à restrição $B' \cdot x = 0$, em que B é uma matriz $m \times n$, $m < n$, e posto igual a m . Esse problema pode ser formulado como:

O lagrangeano do problema acima é

$$L = x' \cdot A \cdot x - \lambda (\|x\|^2 - 1) - 2 \cdot \mu_1 \cdot (b_{11} \cdot x_1 + \dots + b_{1n} \cdot x_n) - \dots - 2 \cdot \mu_m \cdot (b_{m1} \cdot x_1 + \dots + b_{mn} \cdot x_n).$$

Os são apenas uma questão de conveniência para a simplificação a seguir. As condições de 1ª ordem desse problema são:

$$\begin{aligned} 2 \cdot a_{11} \cdot x_1 + \dots + 2 \cdot a_{1n} \cdot x_n - 2 \cdot \lambda \cdot x_1 - 2 \cdot \mu_1 \cdot b_{11} - \dots - 2 \cdot \mu_m \cdot b_{m1} &= 0 \\ 2 \cdot a_{21} \cdot x_1 + \dots + 2 \cdot a_{2n} \cdot x_n - 2 \cdot \lambda \cdot x_2 - 2 \cdot \mu_1 \cdot b_{12} - \dots - 2 \cdot \mu_m \cdot b_{m2} &= 0 \\ \vdots & \\ 2 \cdot a_{m1} \cdot x_1 + \dots + 2 \cdot a_{mn} \cdot x_n - 2 \cdot \lambda \cdot x_n - 2 \cdot \mu_1 \cdot b_{1n} - \dots - 2 \cdot \mu_m \cdot b_{mn} &= 0 \end{aligned}$$

que resulta nas equações abaixo, se simplificarmos e acrescentarmos as restrições $B \cdot x = 0$:

.

A restrição $x_1^2 + \dots + x_n^2 = 1$ empregaremos à frente. Consideremos a notação resumida para o sistema acima: , que diz que $(A - \lambda \cdot I) \cdot x - B \cdot \mu = 0$ e $B' \cdot x = 0$. Certamente, esse sistema tem soluções e μ não nulas, pois e, portanto, existem pontos de máximo e mínimo da forma quadrática quando restrita a e $B' \cdot x = 0$. Vejamos as soluções do sistema acima. Como o posto de B é m , sejam soluções linearmente independentes de $B' \cdot x = 0$, isto é, $x_1, \dots, x_m$. É claro que $x_1, \dots, x_m$ são soluções e satisfazem a restrição . Além disso, $x_1, \dots, x_m$, portanto, $x_1, \dots, x_m$. Na equação $(A - \lambda \cdot I) \cdot x - B \cdot \mu = 0$ $\frac{X'_i}{\|X'_i\|} \cdot (A - \lambda \cdot I) \cdot \frac{X_i}{\|X_i\|} - \frac{X'_i}{\|X'_i\|} \cdot B \cdot \mu = \frac{X'_i}{\|X'_i\|} \cdot (A - \lambda \cdot I) \cdot \frac{X_i}{\|X_i\|} = 0$, isto é, $\mu_i = 0$, ou seja, o valor do multiplicador em cada i é 0. Para obter μ_i para cada i , temos que resolver o sistema . Esses sistemas têm solução, pois o posto de B é m . A conclusão até aqui é que as condições de 1ª ordem do lagrangeano são satisfeitas pelos pontos $x_1, \dots, x_m$. E, como essas soluções são não-nulas, então .

O determinante é um polinômio em de grau $n - m$

.

Nesse ponto, Hancock utiliza uma antiga regra, chamada “Regra de Descartes”¹⁵, que relaciona o número de raízes positivas ou negativas de um polinômio com o número de trocas de sinais nos coeficientes de um polinômio. A Regra de Descartes diz que o número P de raízes positivas de um polinômio $p(\lambda)$, contadas cada

uma com suas multiplicidades, é igual ou menor que o número c de vezes que os coeficientes do polinômio trocam de sinal na sequência dos coeficientes, não contando os coeficientes nulos; se o número p de raízes é menor que o número c de trocas e se c é par, então p é par; se c é ímpar, então p é ímpar. Por exemplo, $p(\lambda)$ tem três trocas de sinal, portanto o número de raízes positivas é 1 ou 3. De fato, as raízes positivas são λ_1 e λ_2 , que tem multiplicidade 1 e 2, respectivamente. Como uma raiz negativa de $p(\lambda)$ é uma raiz positiva de $p(-\lambda)$, para aplicar a regra em relação às raízes negativas, a aplicamos em $p(-\lambda)$. Por exemplo, troca de sinal uma vez, então $p(-\lambda)$ tem, no máximo, uma raiz negativa – e de fato tem uma raiz negativa que é $-\lambda_3$.

Exemplo 40: O polinômio $p(\lambda)$ apresenta duas trocas de sinal, então, necessariamente, $p(\lambda)$ tem duas raízes positivas ou nenhuma raiz positiva. Como $p(-\lambda)$ apresenta uma troca de sinal, então $p(-\lambda)$ tem uma raiz negativa. Portanto, $p(\lambda)$ não tem raízes positivas e uma raiz negativa ou $p(\lambda)$ tem duas raízes positivas e uma raiz negativa. Como $p(\lambda)$ tem uma raiz positiva no intervalo $(0, 1)$, portanto tem outra raiz positiva. Conclusão: $p(\lambda)$ tem duas raízes positivas e uma raiz negativa.

Exemplo 41: Seja a forma quadrática q , cuja matriz é A . O polinômio característico da matriz A é $p(\lambda)$ e, portanto, pela Regra de Descartes, a matriz A tem zero ou dois autovalores positivos e um autovalor negativo. Como a matriz A tem três autovalores reais, então dois autovalores são positivos e um autovalor é negativo e a forma simétrica q é indefinida.

Exercício 14

Use a Regra de Descartes para determinar as definições das formas abaixo:

i); ;

ii); ;

iii); ;

iv); ;

v); ;

vi); .

vii); $p(\lambda) = \lambda^3 + 35 \cdot \lambda^2 + 210 \cdot \lambda$; $p(-\lambda) = -\lambda^3 + 35 \cdot \lambda^2 - 210 \cdot \lambda$.

Respostas: i) positiva-definida; ii) indefinida; iii) positiva-definida; iv) indefinida; v) indefinida; vi) positiva-semidefinida; vii) negativa-semidefinida.

Exemplo 42: Dada a matriz A , então $p_A(\lambda)$ é o polinômio característico de A . Se $a_{11} > 0$ e $a_{22} > 0$, então A tem duas trocas de sinal e nenhuma. Portanto, A é positiva-definida. Se $a_{11} < 0$ e $a_{22} < 0$, então A não tem trocas de sinal e tem duas. Portanto, A é negativa-definida. Digamos que $a_{11} > 0$ e $a_{22} < 0$, então A tem uma troca de sinal e uma. Portanto, A é positiva-semidefinida; se $a_{11} < 0$ e $a_{22} > 0$, então A é negativa-semidefinida. Se $a_{11} = 0$ e $a_{22} = 0$, então A é indefinida.

Para uma matriz simétrica $A, 3 \times 3$, o polinômio característico é:

$$p(\lambda) = |\lambda \cdot I - A| = \lambda^3 - (a_{11} + a_{22} + a_{33}) \cdot \lambda^2 + \left(\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{12} & a_{13} \\ a_{21} & a_{22} \end{vmatrix} \right) \cdot \lambda - |A|.$$

Observe que se $a_{11} > 0$ e $a_{22} > 0$, então A é positiva-definida. Se $a_{11} < 0$ e $a_{22} < 0$, então A é negativa-definida.

Voltando ao polinômio $p_A(\lambda)$. Se todos a_{11}, a_{22}, a_{33} são não-nulos e têm o mesmo sinal, então temos $n-m$ trocas de sinal em $p_A(\lambda)$ e nenhuma troca de sinal em A , logo, as raízes $\lambda_1, \dots, \lambda_n$ são necessariamente positivas e $x' \cdot A \cdot x$ é positiva-definida quando restrita ao $B' \cdot x = 0$, pois o mínimo de $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x = 0$ e $|x| = 1$, é alguma dessas raízes que é positiva. Por outro lado, se todos a_{11}, a_{22}, a_{33} são não-nulos e alternam os sinais, então o polinômio $p_A(\lambda)$ tem $n-m$ trocas de sinais nos coeficientes e $p_A(\lambda)$ nenhuma troca, então as raízes $\lambda_1, \dots, \lambda_n$ são todas negativas e $x' \cdot A \cdot x$ é negativa-

definida quando restrita ao $B \cdot x = 0$, pois o máximo de $x' \cdot A \cdot x$, sujeita às restrições $B' \cdot x = 0$ e $x \geq 0$, é alguma dessas raízes que é negativa.

Logo, o resultado de Hancock propõe que analisemos as trocas de sinais nos coeficientes do polinômio.

Teorema 20 – Teorema de Hancock

Sejam $b_1, \dots, b_n$ os coeficientes do polinômio.

i) Se $b_1, \dots, b_n$ são não-nulos e têm o mesmo sinal, então $x' \cdot A \cdot x$ é positiva-definida quando restrita ao $B' \cdot x = 0$;

ii) Se $b_1, \dots, b_n$ são não-nulos e alternam sinal, então $x' \cdot A \cdot x$ é negativa-definida quando restrita ao $B' \cdot x = 0$.

O resultado obtido por Hancock não é de emprego fácil e por isso procurou-se outros critérios. Por exemplo, o resultado do Hancock foi empregado por Hotelling¹⁶ para obter uma condição que, mais tarde, foi percebida implicando a quase-concavidade das funções utilidades.

9.2 Hotelling e as utilidades

Harold Hotelling (1895-1973) doutorou-se em Matemática em Princeton, no ano de 1924, sob a supervisão de Oswald Veblen, sobrinho do economista. Em 1927, tornou-se Professor Associado no Departamento de Matemática da Universidade de Stanford. Naquele ano revisou o livro do estatístico inglês Ronald Aylmer Fisher (1890-1962), *“Statistical Methods for Research Workers”*, o que foi decisivo no redirecionamento da carreira de Hotelling para a Economia e para a Estatística. Em 1931, Hotelling foi para a Universidade de Columbia, onde substituiu Henry Ludwell Moore (1869 -1958), que foi um economista norte-americano pioneiro no emprego de mínimos quadrados e séries temporais em Economia.

Moore, que foi muito importante no desenvolvimento da Econometria, orientou Henry Schultz (1893-1938), que deu continuidade ao trabalho de Moore na estimativa das funções de demanda de várias *commodities* importantes na época. No ano de sua morte, num acidente automobilístico, Schultz publicou uma das obras mais importantes da literatura econométrica no século 20, o livro "*The Theory and Measurement of Demand*"¹⁷. A presença de Hotelling em Columbia propiciou a cooperação entre ambos e Hotelling dedicou-se a analisar as condições de segunda ordem no problema das demandas individuais, que na época não haviam sido analisadas. As pesquisas de Hotelling constituem a Parte I, "*Theoretical Foundation*", do livro de Schultz. Segundo Louçã¹⁸, a morte de Schultz foi um fator importante para o abandono progressivo da Economia por parte de Hotelling e maior dedicação à Estatística.

Em 1942, Hotelling integrou o *Statistical Research Group* (SRG), um órgão de pesquisa vinculado aos militares e cujos objetivos incluíam estudar problemas relacionados à efetividade de bombardeios antiaéreos, danos em aeronaves sob bombardeio, vulnerabilidade a bombardeios e outros problemas bélicos que envolviam estatística. Hotelling foi o principal cientista do SRG e muitos dos participantes foram seus estudantes. Esse grupo incluiu economistas que, posteriormente, se tornaram famosos, como Milton Friedman. Para Mirowski¹⁹, o SRG propiciou a formação daquilo que mais tarde se chamou a "Escola de Chicago", que trabalhou os problemas da teoria neoclássica do preço sob inspiração dos trabalhos anteriores de Hotelling e Schultz.

Hotelling considera uma função de utilidade , em que são as quantidades dos bens consumidos, sujeita à restrição orçamentária , em que são os preços das *commodities* . Pelo resultado de Hancock, a

condição suficiente para que a cesta x , satisfazendo as condições de primeira ordem do lagrangeano $\mathcal{L}$, seja um máximo local, é que as raízes do polinômio

sejam negativas, pois isso implica que a forma quadrática $\nabla^2 \mathcal{L}$ seja negativa-definida no subespaço determinado pelo hiperplano $\nabla \mathcal{L} = 0$. Em qualquer dessas raízes é tal que

.

Vejamos o caso $n=2$, que não apresenta dificuldade para o cálculo.

. Como $\lambda > 0$, então x é condição suficiente para ser um mínimo local.

Vejamos o caso $n \geq 3$. O cálculo dos determinantes para $n \geq 3$ é árduo. No caso,

Como as duas raízes são negativas, então $\lambda > 0$.

Podemos imaginar três situações. Na primeira, fixamos a quantidade ótima x^* e propomos o problema:

Obviamente, a solução desse problema são as quantidades x^* e λ^* . Nesse problema, para que as raízes de

sejam negativas, pois isso implica que a forma quadrática $\nabla^2 \mathcal{L}$ é negativa-definida no subespaço determinado pelo hiperplano $\nabla \mathcal{L} = 0$, a

exigência é

.

As outras duas situações são os problemas:

$$\begin{aligned} & \max_{q_1^*, q_2^*, q_3^*} \phi(q_1^*, q_2^*, q_3^*) \\ \text{e} \quad & \text{sa} \quad p_2 \cdot q_2 + p_3 \cdot q_3 = m - p_1 \cdot q_1^*, \\ & q_2 > 0, q_3 > 0. \end{aligned}$$

que igualmente são tais que

e

implica que a forma quadrática seja negativa-definida nos hiperplanos e , respectivamente. Ora, estes determinantes e

são os mesmos que aparecem no polinômio . Se passamos a utilizar a notação

,

o polinômio acima se expressa como . E, repetindo, se e , então é um máximo local para sujeita às restrições e . Observe que a exigência é , ou seja, os determinantes de ordem par sejam positivos e os de ordem ímpares sejam negativos.

No caso geral,

E, consistentemente, devemos concluir que se , então é um máximo local para sujeita às restrições e .

Estudando diretamente as condições de primeira e segunda ordens para o problema do consumidor, Eugen Slutsky (1880-1948) estabeleceu o mesmo teorema num artigo publicado em 1915, “*Sulla Teoria del Bilancio del Consumatore*”²⁰. Redescoberto, esse artigo

mereceu uma resenha detalhada do inglês Roy G. D. Allen²¹ (1906-1983).

No livro “*Capital and Value*”, de John Hicks (1904-1989), de 1939, a referência é o artigo de Slutsky²². No apêndice matemático do livro, Hicks observa que o resultado pode ser obtido substituindo a variável p por m e tratando o problema como um problema de otimização não-condicionada nas variáveis x . Essa substituição transforma o problema de avaliar a condição de segunda ordem do ponto crítico do lagrangeano num problema de avaliar as condições de segunda ordem do ponto crítico da função L . Esse método foi posteriormente adotado por Arrow²³. Tais condições são fáceis de serem obtidas num problema com duas variáveis (q_1, q_2) , mas não são tão simples para mais variáveis e exigem certa engenharia algébrica.

9.3 A demonstração de Mann

Henry B. Mann²⁴(1905-2000) procurou um obter um resultado mais simples que o de Hancock e, para isso, recorreu ao critério de Jacobi para obter uma condição necessária e suficiente para uma forma ser positiva-definida ou negativa-definida num subespaço. O teorema de Mann é muito citado e foi reproduzido no volume I do clássico “*Readings in Mathematical Economics*”²⁵. Mann foi o primeiro a obter uma condição necessária e suficiente para uma forma ser positiva-definida ou negativa-definida num subespaço $B' \cdot x = 0$, tal que B' é uma matriz $p \times n$, através dos determinantes menores principais líderes da matriz $\begin{vmatrix} A & B' \\ B' & 0 \end{vmatrix}$.

O artigo de Mann está relacionado aos interesses de Hotelling. Henry Bertold Mann foi um matemático dedicado à Teoria dos Números, mas publicou alguns trabalhos em Estatística. Em 1942, Mann trabalhou com Abraham Wald (1902-1950) e Hotelling na

Universidade de Columbia, no Departamento de Economia, que na época era dirigido por Hotelling²⁶. O artigo sobre formas quadráticas resulta provavelmente do interesse de Hotelling no assunto. Mais adiante, observaremos que foi Hotelling quem apresentou o artigo de Mann para Debreu.

Mann não tratou de formas semipositivas e seminegativas e essa situação ligeiramente mais geral foi tratada por Debreu no já referido artigo do autor.

Mann provou o resultado que obteve através de uma troca de coordenadas convenientes para expressar a forma quadrática. Entretanto, a exposição de Mann não é clara, por isso preferimos uma abordagem mais acessível.

Dada a forma quadrática $x' \cdot A \cdot x = (x_1 \ \dots \ x_n) \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ sujeita às restrições

$$B' \cdot x = \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, p < n, \text{ isto é, } x_1 = 0, x_2 = 0, \dots, x_p = 0. \text{ A matriz da forma}$$

quadrática restrita ao subespaço $\{(0, \dots, 0, x_{p+1}, \dots, x_n)\}$ é $\begin{pmatrix} a_{(p+1)(p+1)} & \dots & a_{(p+1)n} \\ \vdots & \ddots & \vdots \\ a_{n(p+1)} & \dots & a_{nn} \end{pmatrix}$, portanto se a forma é positiva ou negativa-definida, determina-se pelos sinais dos

determinantes $|a_{(p+1)(p+1)}|, |a_{(p+1)(p+1)} \ a_{(p+1)(p+2)}|, \dots, |a_{(p+1)(p+1)} \ \dots \ a_{(p+1)n}|$. Observe que esses

$$\text{determinantes são tamb} |h_{p+2}| = \begin{vmatrix} a_{11} & \dots & \dots & a_{1(p+2)} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \dots & 0 & 1 \\ a_{(p+1)1} & \dots & a_{(p+1)(p+1)} & a_{(p+1)(p+2)} & 0 & \dots & 0 \\ a_{(p+2)1} & \dots & a_{(p+2)(p+1)} & a_{(p+2)(p+2)} & 0 & \dots & 0 \\ 1 & \vdots & 0 & 0 & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 \end{vmatrix} = (-1)^p \cdot \begin{vmatrix} a_{(p+1)(p+1)} & a_{(p+1)(p+2)} \\ a_{(p+2)(p+1)} & a_{(p+2)(p+2)} \end{vmatrix} \text{ é}$$

$$\text{definidos por } |h_{p+1}| = \begin{vmatrix} a_{11} & \dots & a_{1(p+1)} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ \vdots & \ddots & \vdots & 0 & \dots & 1 \\ a_{(p+1)1} & \dots & a_{(p+1)(p+1)} & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & \dots & 0 \end{vmatrix} = (-1)^p \cdot a_{(p+1)(p+1)} = (-1)^p \cdot |a_{(p+1)(p+1)}|, \dots,$$

$$\left| h_{p+(n-p)} \right| = \begin{vmatrix} a_{11} & \cdots & a_{1(p+1)} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \ddots & \vdots & \cdots & 0 & \cdots & 1 \\ a_{(p+1)1} & \cdots & a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} & 0 & \cdots & 0 \\ \vdots & & \vdots & & \ddots & \vdots & & \vdots \\ 0 & \vdots & a_{n(p+1)} & \cdots & a_{nn} & 0 & \cdots & 0 \\ 1 & \vdots & 0 & \vdots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 \end{vmatrix} = (-1)^p \cdot \begin{vmatrix} a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} \\ \vdots & \ddots & \vdots \\ a_{n(p+1)} & \cdots & a_{nn} \end{vmatrix}$$

Portanto, concluímos que $|a_{(p+1)(p+1)}| = (-1)^p \cdot |h_{p+1}|$, $\begin{vmatrix} a_{(p+1)(p+1)} & a_{(p+1)(p+2)} \\ a_{(p+2)(p+1)} & a_{(p+2)(p+2)} \end{vmatrix} = (-1)^p \cdot |h_{p+2}|$, ..., $\begin{vmatrix} a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} \\ \vdots & \ddots & \vdots \\ a_{n(p+1)} & \cdots & a_{nn} \end{vmatrix} = (-1)^p \cdot |h_{p+(n-p)}|$. Logo, a forma $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x = 0$, é positiva-definida se, e só se, $(-1)^p \cdot |h_{p+1}| = |a_{(p+1)(p+1)}| > 0$, $(-1)^p \cdot |h_{p+2}| = \begin{vmatrix} a_{(p+1)(p+1)} & a_{(p+1)(p+2)} \\ a_{(p+2)(p+1)} & a_{(p+2)(p+2)} \end{vmatrix} > 0, \dots$, $(-1)^p \cdot |h_n| = \begin{vmatrix} a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} \\ \vdots & \ddots & \vdots \\ a_{n(p+1)} & \cdots & a_{nn} \end{vmatrix} > 0$; a forma $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x = 0$, é negativa-definida se, e só se, $(-1)^1 \cdot |a_{(p+1)(p+1)}| > 0$, $(-1)^2 \cdot \begin{vmatrix} a_{(p+1)(p+1)} & a_{(p+1)(p+2)} \\ a_{(p+2)(p+1)} & a_{(p+2)(p+2)} \end{vmatrix} > 0, \dots$, $(-1)^{n-k} \cdot \begin{vmatrix} a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} \\ \vdots & \ddots & \vdots \\ a_{n(p+1)} & \cdots & a_{nn} \end{vmatrix} > 0$, ou seja, $(-1)^1 \cdot |a_{(p+1)(p+1)}| = (-1)^{p+1} \cdot |h_1| > 0$, $(-1)^2 \cdot \begin{vmatrix} a_{(p+1)(p+1)} & a_{(p+1)(p+2)} \\ a_{(p+2)(p+1)} & a_{(p+2)(p+2)} \end{vmatrix} = (-1)^{p+2} \cdot |h_2| > 0$, ..., $(-1)^{n-p} \cdot \begin{vmatrix} a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} \\ \vdots & \ddots & \vdots \\ a_{n(p+1)} & \cdots & a_{nn} \end{vmatrix} = (-1)^{p+(n-p)} \cdot |h_n| > 0$.

Teorema 21

i) A forma $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x = 0$, é negativa-definida se, e só se, $(-1)^{p+1} \cdot |h_{p+1}| > 0$, $(-1)^{p+2} \cdot |h_{p+2}| > 0$, $(-1)^{p+3} \cdot |h_{p+3}| > 0$, ..., $(-1)^{n+(n-p)} \cdot |h_n| > 0$;

ii) A forma $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x = 0$, é positiva-definida se, e só se, $(-1)^p \cdot |h_{p+1}| > 0$, $(-1)^p \cdot |h_{p+2}| > 0$, $(-1)^p \cdot |h_{p+3}| > 0$, ..., $(-1)^p \cdot |h_n| > 0$.

Mostramos esse resultado para o caso particular em que B' é a

matriz $p \times n$ $B' = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \underset{\uparrow}{1} & 0 & \cdots & 0 \end{pmatrix}, p < n$.

Na argumentação anterior, empregamos o resultado

$$\begin{vmatrix} 0_{m \times n} & A_{m \times m} \\ B_{n \times n} & 0_{n \times m} \end{vmatrix} = (-1)^{m \cdot n} \cdot \begin{vmatrix} B_{n \times n} & 0_{n \times m} \\ 0_{m \times n} & A_{m \times m} \end{vmatrix} = (-1)^{m \cdot n} \cdot |A| \cdot |B|, \text{ em que } A_{m \times m} \text{ e } B_{n \times n} \text{ são matrizes } m \times m \text{ e } n \times n,$$

respectivamente. Este resultado pode ser facilmente estendido para

obter que $\begin{vmatrix} 0 & 0 & A_{m \times m} \\ 0 & B_{n \times n} & 0 \\ C_{p \times p} & 0 & 0 \end{vmatrix} = (-1)^{m \cdot q + n \cdot q + m \cdot n} \cdot |A| \cdot |B| \cdot |C|$. A partir deste resultado, verificamos

$$\text{que } |h_{p+k}| = \begin{vmatrix} a_{11} & \cdots & a_{1(p+1)} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots & \cdots & 0 & \cdots & 1 \\ a_{(p+1)1} & \cdots & a_{(p+1)(p+1)} & \cdots & a_{(p+1)n} & 0 & \cdots & 0 \\ \vdots & & \vdots & & \ddots & \vdots & & \vdots \\ a_{(p+k)1} & \cdots & a_{(p+k)(p+1)} & \cdots & a_{(p+k)(p+k)} & 0 & \cdots & 0 \\ 1 & \cdots & 0 & \cdots & 0 & \vdots & \vdots & \vdots \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1 & 0 & \cdots & 0 & \cdots & \cdots & 0 \end{vmatrix} = (-1)^p \cdot \begin{vmatrix} a_{(p+1)(p+1)} & \cdots & a_{(p+1)(p+k)} \\ \vdots & \ddots & \vdots \\ a_{(p+k)(p+1)} & \cdots & a_{(p+k)(p+k)} \end{vmatrix}, \text{ para } k=1, \dots, n-p, \text{ pois}$$

$$\begin{vmatrix} a_{11} & \cdots & a_{1(p+1)} & \cdots & a_{1(p+k)} & 1 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & \vdots & & \vdots \\ a_{p1} & \cdots & a_{p(p+1)} & \cdots & a_{p(p+k)} & 0 & \cdots & 1 \\ a_{(p+1)1} & \cdots & a_{(p+1)(p+1)} & \cdots & a_{(p+1)(p+k)} & 0 & \cdots & 0 \\ \vdots & & \vdots & & \ddots & \vdots & & \vdots \\ a_{(p+k)1} & \cdots & a_{(p+k)(p+1)} & \cdots & a_{(p+k)(p+k)} & 0 & \cdots & 0 \\ 1 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{vmatrix} =$$

$$\begin{vmatrix} 0 & \cdots & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \vdots & \ddots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ \hline 0 & \cdots & 0 & a_{(p+1)(p+1)} & \cdots & a_{(p+1)(p+k)} & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{(p+k)(p+1)} & \cdots & a_{(p+k)(p+k)} & \cdots & 0 \\ \hline 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{vmatrix}$$

Exemplo 43: Seja a forma quadrática $q(x, y, z) = 3 \cdot x^2 - 2x \cdot y - y^2 + 4 \cdot x \cdot z - 2 \cdot y \cdot z - 2 \cdot z^2$, cuja matriz é $\begin{pmatrix} 3 & -1 & 2 \\ -1 & -1 & -1 \\ 2 & -1 & -2 \end{pmatrix}$. A forma é indefinida, pois, por exemplo, os menores principais de ordem 1 não têm o mesmo sinal.

No plano y, z , isto é, $x=0$, $(-1)^2 \cdot \begin{vmatrix} 3 & -1 & 1 \\ -1 & -1 & 0 \\ 1 & 0 & 0 \end{vmatrix} = 1 > 0$ e $(-1)^3 \cdot \begin{vmatrix} 3 & -1 & 2 & 1 \\ -1 & -1 & -1 & 0 \\ 2 & -1 & -2 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} = 1 > 0$, logo q é negativa-definida no plano y, z . Observe que nesse plano, $q(0, y, z) = -y^2 - 2 \cdot y \cdot z - 2 \cdot z^2$.

Exemplo 44: Seja a forma quadrática $q(x, y, z) = -x^2 - 2x \cdot y + 3 \cdot y^2 - 6 \cdot x \cdot z - 4 \cdot y \cdot z + 5 \cdot z^2$, cuja matriz é $\begin{pmatrix} -1 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 5 \end{pmatrix}$. A forma é indefinida. No plano y, z , isto é, $x=0$,

$(-1)^1 \cdot \begin{vmatrix} -1 & -1 & 1 \\ -1 & 3 & 0 \\ 1 & 0 & 0 \end{vmatrix} = 3 > 0$ e $(-1)^1 \cdot \begin{vmatrix} -1 & -1 & -3 & 1 \\ -1 & 3 & -2 & 0 \\ -3 & -2 & 5 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} = 11 > 0$, logo q é positiva-definida no plano y, z .

Observe que nesse plano, $q(0, y, z) = 3 \cdot y^2 - 4 \cdot y \cdot z + 5 \cdot z^2$.

Obviamente, o objetivo é mostrar que o resultado é válido para uma matriz $B' = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pn} \end{pmatrix}$, $p < n$. Consideremos a submatriz $B'_{p \times p} = \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pp} \end{pmatrix}$ e suponhamos que $|B'_{p \times p}| = \begin{vmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pp} \end{vmatrix} \neq 0$.

Consideremos a troca de sistema de coordenadas:

$$\begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 + \cdots + b_{1n}x_n \\ &\vdots \\ y_p &= b_{p1}x_1 + b_{p2}x_2 + \cdots + b_{pn}x_n \\ y_{p+1} &= x_{p+1} \\ &\vdots \\ y_n &= x_n, \end{aligned}$$

ou seja,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_p \\ y_{p+1} \\ \vdots \\ y_n \end{pmatrix} = \left(\begin{array}{ccc|ccc} b_{11} & \cdots & b_{1p} & b_{1(p+1)} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{p1} & \cdots & b_{pp} & b_{p(p+1)} & \cdots & b_{pn} \\ \hline 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{array} \right) \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ x_{p+1} \\ \vdots \\ x_n \end{pmatrix}.$$

Em termos matriciais, $y = \begin{pmatrix} B'_{p \times p} & B'_{p \times (n-p)} \\ 0 & I_{(n-p) \times (n-p)} \end{pmatrix} \cdot x$. Como a forma quadrática f é $y' \cdot \bar{A} \cdot y = x' \cdot \begin{pmatrix} B'_{p \times p} & B'_{p \times (n-p)} \\ 0 & I_{(n-p) \times (n-p)} \end{pmatrix} \cdot \bar{A} \cdot \begin{pmatrix} B'_{p \times p} & B'_{p \times (n-p)} \\ 0 & I_{(n-p) \times (n-p)} \end{pmatrix} \cdot x = x' \cdot A \cdot x$, então $A = \begin{pmatrix} B'_{p \times p} & B'_{p \times (n-p)} \\ 0 & I_{(n-p) \times (n-p)} \end{pmatrix} \cdot \bar{A} \cdot \begin{pmatrix} B'_{p \times p} & B'_{p \times (n-p)} \\ 0 & I_{(n-p) \times (n-p)} \end{pmatrix}$.

Observe que a restrição $B' \cdot x = 0$, em termos das coordenadas x , se expressa como $y_1 = 0, \dots, y_p = 0$, em termos das coordenadas y . Portanto, se $\bar{A} = \begin{pmatrix} \bar{a}_{11} & \cdots & \bar{a}_{1n} \\ \vdots & \ddots & \vdots \\ \bar{a}_{n1} & \vdots & \bar{a}_{nn} \end{pmatrix}$, a definição da forma f se determina através dos determinantes

$$|\bar{h}_{p+1}| = \begin{vmatrix} \bar{a}_{11} & \cdots & \bar{a}_{1(p+1)} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ \vdots & \ddots & \vdots & 0 & \cdots & 1 \\ \hline \bar{a}_{(p+1)1} & \cdots & \bar{a}_{(p+1)(p+1)} & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & \cdots & 0 \end{vmatrix} = (-1)^p \cdot \bar{a}_{(p+1)(p+1)} = (-1)^p \cdot |\bar{a}_{(p+1)(p+1)}|$$

$$|\bar{h}_{p+2}| = \begin{vmatrix} \bar{a}_{11} & \cdots & \cdots & \bar{a}_{1(p+2)} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \cdots & 0 & 1 \\ \bar{a}_{(p+1)1} & \cdots & \bar{a}_{(p+1)(p+1)} & \bar{a}_{(p+1)(p+2)} & 0 & \cdots & 0 \\ \bar{a}_{(p+2)1} & \cdots & \bar{a}_{(p+2)(p+1)} & \bar{a}_{(p+2)(p+2)} & 0 & \cdots & 0 \\ 1 & \vdots & 0 & 0 & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 \end{vmatrix} = (-1)^p \cdot \begin{vmatrix} \bar{a}_{(p+1)(p+1)} & \bar{a}_{(p+1)(p+2)} \\ \bar{a}_{(p+2)(p+1)} & \bar{a}_{(p+2)(p+2)} \end{vmatrix}, \dots,$$

$$|\bar{h}_n| = \begin{vmatrix} \bar{a}_{11} & \cdots & \bar{a}_{1(p+1)} & \cdots & \bar{a}_{1n} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \cdots & 0 & \cdots & 1 \\ \bar{a}_{(p+1)1} & \cdots & \bar{a}_{(p+1)(p+1)} & \cdots & \bar{a}_{(p+1)n} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \bar{a}_{n(p+1)} & \cdots & \bar{a}_{nn} & 0 & \cdots & 0 \\ 1 & \vdots & 0 & \vdots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 \end{vmatrix} = (-1)^p \cdot \begin{vmatrix} \bar{a}_{(p+1)(p+1)} & \cdots & \bar{a}_{(p+1)n} \\ \vdots & \ddots & \vdots \\ \bar{a}_{n(p+1)} & \cdots & \bar{a}_{nn} \end{vmatrix}.$$

Portanto, o problema se reduz à análise da relação entre os determinantes

$$|\bar{h}_{p+i}| = \begin{vmatrix} \bar{a}_{11} & \cdots & \cdots & \bar{a}_{1(p+i)} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \bar{a}_{pp} & \vdots & \vdots & \cdots & 0 & 1 \\ \cdots & \cdots & \ddots & \cdots & 0 & \cdots & 0 \\ \bar{a}_{(p+i)1} & \cdots & \cdots & \bar{a}_{(p+i)(p+i)} & 0 & \cdots & 0 \\ 1 & \vdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 \end{vmatrix} = (-1)^p \cdot \begin{vmatrix} \bar{a}_{(p+1)(p+1)} & \cdots & \bar{a}_{(p+1)(p+i)} \\ \vdots & \ddots & \vdots \\ \bar{a}_{(p+i)(p+1)} & \cdots & \bar{a}_{(p+i)(p+i)} \end{vmatrix} \quad e$$

$$|h_{p+i}| = \begin{vmatrix} a_{11} & \cdots & \cdots & a_{1(p+i)} & b_{11} & \cdots & b_{p1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{(p+i)1} & \cdots & \cdots & a_{(p+i)(p+i)} & b_{(p+i)1} & \cdots & b_{p(p+i)} \\ b_{11} & \cdots & \cdots & b_{(p+i)1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{p1} & \cdots & \cdots & b_{p(p+i)} & 0 & \cdots & 0 \end{vmatrix}, \text{ para } i=1, \dots, n-p.$$

Lema 3

$$|h_{p+i}| = |B'_{p \times p}|^2 \cdot |\bar{h}_{p+i}| = (-1)^p \cdot |B'_{p \times p}|^2 \cdot \begin{vmatrix} \bar{a}_{(p+1)(p+1)} & \cdots & \bar{a}_{(p+1)(p+i)} \\ \vdots & \ddots & \vdots \\ \bar{a}_{(p+i)(p+1)} & \cdots & \bar{a}_{(p+i)(p+i)} \end{vmatrix}, i=1, \dots, n-p.$$

Prova: primeiramente observe que, como $A = \left(\begin{array}{c|c} B'_{p \times p} & B'_{p \times (n-p)} \\ \hline 0 & I_{(n-p) \times (n-p)} \end{array} \right)' \cdot \bar{A} \cdot \left(\begin{array}{c|c} B'_{p \times p} & B'_{p \times (n-p)} \\ \hline 0 & I_{(n-p) \times (n-p)} \end{array} \right)$, então $|h_n| = |B'_{p \times p}|^2 \cdot |\bar{h}_n|$. Se eliminamos a variável x_n , ou seja, fazemos $x_n = 0$, estamos com a forma quadrática f_{n-1} , cuja matriz em $\mathbb{R}^{n-1}$ é $A = \begin{pmatrix} a_{11} & \cdots & a_{1(n-1)} \\ \vdots & \ddots & \vdots \\ a_{(n-1)1} & \cdots & a_{(n-1)(n-1)} \end{pmatrix}$

e a restrição $\begin{cases} b_{11}x_1 + b_{12} \cdot x_2 + \cdots + b_{1n} \cdot x_{n-1} = 0 \\ b_{p1}x_1 + b_{p2} \cdot x_2 + \cdots + b_{pn} \cdot x_{n-1} = 0. \end{cases}$

Nesse caso, fazemos a troca de coordenadas

$$\begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 + \dots + b_{1(n-1)}x_{n-1} \\ &\vdots \\ y_p &= b_{p1}x_1 + b_{p2}x_2 + \dots + b_{p(n-1)}x_{n-1} \\ y_{p+1} &= x_{p+1} \\ &\vdots \\ y_{n-1} &= x_{n-1}, \end{aligned}$$

Como anteriormente, obtemos que $|h_{n-1}| = |B'_{p \times p}|^2 |\bar{h}_{n-1}|$. Podemos prosseguir fazendo até $x_n=0, x_{n-1}=0, \dots, x_{p+2}=0$.

Através da expressão do lema, observamos que $|h_{p+i}|$ e $|\bar{h}_{p+i}|$ têm o mesmo sinal e, como a definição da forma f em $B' \cdot x=0$ depende dos sinais dos $|\bar{h}_{p+i}|$, que são os mesmos dos $|h_{p+i}|$, então podemos estabelecer o resultado abaixo.

Teorema 22 – Teorema de Mann

i) A forma $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x=0, |B'_{p \times p}| \neq 0$, é negativa-definida se, e só se, $(-1)^{p+1} \cdot |h_{p+1}| > 0, (-1)^{p+2} \cdot |h_{p+2}| > 0, (-1)^{p+3} \cdot |h_{p+3}| > 0, \dots, (-1)^{n-(n-p)} \cdot |h_n| > 0$;

ii) A forma $x' \cdot A \cdot x$, sujeita à restrição $B' \cdot x=0, |B'_{p \times p}| \neq 0$, é positiva-definida se, e só se, $(-1)^p \cdot |h_{p+1}| > 0, (-1)^p \cdot |h_{p+2}| > 0, (-1)^p \cdot |h_{p+3}| > 0, \dots, (-1)^p \cdot |h_n| > 0$.

Exemplo 45: Já vimos que a forma quadrática $q(x, y, z) = x^2 - 6 \cdot x \cdot y + 5 \cdot y^2 - 2 \cdot x \cdot z + 2 \cdot y \cdot z + 5 \cdot z^2$ é indefinida. A matriz dessa forma é $\begin{pmatrix} 1 & -3 & -1 \\ -3 & 5 & 1 \\ -1 & 1 & 5 \end{pmatrix}$. Observe que $\begin{vmatrix} 1 & -3 \\ -3 & 5 \end{vmatrix} = -4$, o que assegura que a forma é indefinida. $(-1)^1 \cdot \begin{vmatrix} 1 & -3 & 2 \\ -3 & 5 & -1 \\ 2 & -1 & 0 \end{vmatrix} = 9 > 0$ e $(-1)^1 \cdot \begin{vmatrix} -1 & -3 & 2 & 2 \\ -3 & 5 & -1 & -1 \\ -3 & -2 & 5 & 2 \\ 2 & -1 & 2 & 0 \end{vmatrix} = 73 > 0$, logo q é positiva-definida no plano definido por $2 \cdot x - y + 2 \cdot z = 0$.

9.4 O apêndice matemático de Samuelson

Na parte V, “*Quadratic Forms Definite under Constraints*”, do apêndice matemático do livro “*Foundations of Economic Analysis*”,

Samuelson apresenta uma versão menos elaborada do Teorema de Mann, tal que a restrição é dada por $B' \cdot x = 0$, B' é uma matriz $m \times n$ ²⁷.

Samuelson esboça uma demonstração do resultado combinando parte da argumentação de Hancock com o fato que uma forma quadrática é diagonalizável e com a expressão do polinômio

$p(\lambda) = \begin{vmatrix} A - \lambda \cdot I & B \\ B' & 0_{m \times m} \end{vmatrix} = B_0 \cdot (-\lambda)^{n-m} + B_1 \cdot (-\lambda)^{n-m-1} + \dots + B_{n-m}$ em termos dos determinantes menores

principais da matriz $\begin{vmatrix} A & B \\ B' & 0_{m \times m} \end{vmatrix}$. $p(\lambda) = \begin{vmatrix} A - \lambda \cdot I & B \\ B' & 0_{m \times m} \end{vmatrix} = (-1)^m \cdot |B' \cdot B| \cdot (-\lambda)^{n-m} + (\sum |B_{m+1}|) \cdot (-\lambda)^{n-(m+1)} + (\sum |B_{m+2}|) \cdot (-\lambda)^{n-(m+2)} + \dots + (\sum |B_{m+(n-1)}|)$.

Na expressão acima, $\sum |B_{m+k}|$ é a soma dos determinantes menores principais da matriz $\begin{vmatrix} A & B \\ B' & 0_{p \times p} \end{vmatrix}$. Para o menor principal líder de ordem $m+k$, usamos a notação $|h_{m+k}|$ no parágrafo anterior. Samuelson não fornece uma demonstração para a expressão acima de $p(\lambda)$ que é, sem dúvida, a parte mais trabalhosa para uma demonstração rigorosa. Samuelson observa que as raízes do polinômio $p(\lambda) = \begin{vmatrix} A - \lambda \cdot I & B \\ B' & 0_{m \times m} \end{vmatrix}$ são os autovalores da forma, quando restrita ao subespaço $B' \cdot x = 0$, usando a Regra de Descartes e a expressão dos coeficientes dos polinômios, através das somas dos menores orlados $\sum |B_{m+k}|$, conclui o Teorema de Mann.

É difícil saber se Samuelson teria alguma prioridade em relação ao trabalho de Mann. No prefácio ao "*Foundations*", ele diz que a maior parte do conteúdo do livro estava pronto desde 1937, que uma primeira versão do livro foi apresentada em 1941, mas que, no entanto, a 2ª Guerra retardou a publicação até 1947. Não sabemos como foram as versões anteriores do apêndice matemático sem uma pesquisa direta nos manuscritos do autor.

O interesse de Samuelson está relacionado à sua tentativa de determinar proposições empiricamente significativas em Teoria Econômica. Afora essa última consideração, do ponto de vista estritamente matemático, é claro que o problema era estabelecer uma demonstração rigorosa do resultado, como Mann conseguiu em 1943. O resultado em si era conhecido. Isso explica porque a análise

de Mann foi publicada como uma “Note” do “*American Mathematical Monthly*”, ou seja, uma exposição breve de uma nova demonstração ou a demonstração de um resultado conhecido. Os matemáticos não estavam preocupados em estabelecer o resultado com rigor. Esse não era um trabalho na fronteira da pesquisa em Matemática.

9.5 A demonstração de Debreu

Gerard Debreu (1921-2004) foi um matemático francês naturalizado norte-americano. Aparentemente, a decisão de dedicar-se à teoria econômica foi determinada pelo livro “*À La Recherche d’une Discipline Économique*”, de Maurice Allais (1911-2010), publicado em 1943. Em 1950, Debreu tornou-se pesquisador da Cowles Foundation e, em 1955, professor em Yale. Em 1956, uma versão do livro “*Theory of Value, an Axiomatic Analysis of Economic Equilibrium*”, que seria publicado em 1959 pela Fundação Cowles, foi apresentada como tese de doutorado na Universidade de Paris. Em 1962, Debreu tornou-se professor na Universidade da Califórnia, em Berkeley. Outra influência importante sobre Debreu foi o livro “*Theory of Games and Economic Behavior*”, de von Neumann e Morgenstern, de 1944.

Em 1952, Debreu publicou o artigo “*Definite and Semidefinite Quadratic Forms*”²⁸, tema desta seção. Em nota de rodapé, Debreu afirma que o trabalho resultou da insistência de Jacob Marschak (1898-1977), que foi diretor de pesquisa da Cowles Commission de janeiro de 1943 até 1948. Na mesma nota, Debreu esclarece que Hotelling o alertou para a existência do artigo de H. Mann.

Toda a argumentação de Debreu se baseia no seguinte resultado, que ele demonstra na página 296 do referido artigo: seja A simétrica $n \times n$ e B matriz $n \times m$, $m < n$. Então, i) $x' \cdot A \cdot x > 0$, $\forall x \neq 0$, tal que $B' \cdot x = 0$ se, e somente se, existe $\bar{\lambda} \in \mathbb{R}$, tal que a forma quadrática $(x' \cdot A \cdot x + \bar{\lambda} \cdot x' \cdot B \cdot B' \cdot x)$ é positiva-definida

no $\mathbb{R}^n$ para $\lambda > \bar{\lambda}$; ii) $x' \cdot A \cdot x < 0, \forall x \neq 0$, tal que $B' \cdot x = 0$ se, e somente se, existe $\bar{\lambda} \in \mathbb{R}$, tal que a forma quadrática $(x' \cdot A \cdot x + \lambda \cdot x' \cdot B \cdot B' \cdot x)$ é negativa-definida no $\mathbb{R}^n$, para $\lambda < \bar{\lambda}$ ²⁹

A partir desse teorema, com o auxílio de um lema sobre determinantes, e definindo os hessianos $|H_r| = \begin{pmatrix} A_r & B_r \\ B_r' & 0_m \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1r} & b_{11} & \dots & b_{m1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{r1} & \dots & a_{rr} & b_{1r} & \dots & b_{mr} \\ b_{11} & \dots & b_{1r} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mr} & 0 & \dots & 0 \end{pmatrix}, \dots,$

$r = m+1, \dots, n$, Debreu estabelece o Teorema 4, na página 297: se A é matriz simétrica $n \times n$, $|B_m| \neq 0$. Então, i) $x' \cdot A \cdot x > 0, \forall x \neq 0$, tal que $B' \cdot x = 0$ se, e somente se, $(-1)^m \cdot \begin{vmatrix} A_r & B_r \\ B_r' & 0_m \end{vmatrix} > 0, r = m+1, \dots, n$; ii) $x' \cdot A \cdot x < 0, \forall x \neq 0$, tal que $B' \cdot x = 0$ se, e somente se, $(-1)^r \cdot \begin{vmatrix} A_r & B_r \\ B_r' & 0_m \end{vmatrix} > 0, r = m+1, \dots, n$ ³⁰ Esse resultado é o mesmo que Mann obteve em 1943.

Debreu esquivou-se de utilizar o polinômio $p(\lambda) = \begin{vmatrix} A - \lambda \cdot I & B \\ B' & 0_{m \times m} \end{vmatrix}$ para obter os menores principais da matriz $\begin{vmatrix} A & B \\ B' & 0_{m \times m} \end{vmatrix}$, como está implícito na demonstração de Hancock e explícito na demonstração do Hotelling ou na sugestão do Samuelson. Debreu substituiu o polinômio acima pelos polinômios cujos coeficientes da maior potência são os menores principais líderes de ordem r : $0 < |A_r + \lambda \cdot B_{rm} \cdot B_{rm}'| = (-1)^m \cdot \begin{vmatrix} A_r & B_r \\ B_r' & 0 \end{vmatrix} \cdot \lambda^m + \dots$.

Depois de estabelecer o Teorema 6 na página 29: i) $x' \cdot A \cdot x \geq 0, \forall x$ tal que $B' \cdot x = 0$, se, e somente se, $x' \cdot A \cdot x + \alpha \cdot x' \cdot x > 0, \forall \alpha > 0$ e $\forall x \neq 0$ tal que $B' \cdot x = 0$; ii) $x' \cdot A \cdot x \leq 0, \forall x$ tal que $B' \cdot x = 0$, se, e somente se, $x' \cdot A \cdot x + \alpha \cdot x' \cdot x < 0, \forall \alpha < 0$ e $\forall x \neq 0$ tal que $B' \cdot x = 0$ e usando permutações para obter os demais menores principais, estabelece o resultado final nos teoremas 9 e 10: i) seja A simétrica $n \times n$, $|B_m| \neq 0$. $x' \cdot A \cdot x \geq 0, \forall x$ tal que $B' \cdot x = 0$, se, e somente se, $(-1)^m \cdot \begin{vmatrix} A_r^\pi & B_r^\pi \\ B_r^{\pi'} & 0_m \end{vmatrix} \geq 0, r = m+1, \dots, n$ e $\forall \pi$; ii) Seja A simétrica $n \times n$, $|B_m| \neq 0$. $x' \cdot A \cdot x \leq 0, \forall x$ tal que $B' \cdot x = 0$, se, e somente se, $(-1)^r \cdot \begin{vmatrix} A_r^\pi & B_r^\pi \\ B_r^{\pi'} & 0_m \end{vmatrix} \geq 0, r = m+1, \dots, n$ e $\forall \pi$ ³¹

Portanto, o artigo de Debreu, que hoje é a referência no assunto, é um conjunto de resultados sobre as definições de formas quadráticas restritas ou não, cobrindo um período de dois séculos, para os quais

o autor forneceu demonstrações rigorosas, sendo que, em relação às definições das formas em subespaços, apresentou uma argumentação inovadora.

Outro matemático também realizou trabalho semelhante. O enunciado do teorema de Mann está no parágrafo 7.98 do livro *“Linear Algebra”*, do matemático russo Georgi E. Shilov (1917-1975). Shilov remete a um artigo escrito em russo, publicado em 1954, de autoria de um matemático chamado R. Ya. Shostak³².

9. 6 O teorema de Arrow-Enthoven

Kenneth Arrow (1921-2017) concluiu o doutorado na Universidade de Columbia em 1950, sob a supervisão de Hotelling. Arrow foi associado de pesquisa na Comissão Cowles de abril de 1947 até junho de 1949. Em 1948, foi para Universidade de Stanford, onde permaneceu até aposentar-se. Os trabalhos mais conhecidos de Arrow são a monografia *“Social Choice and Individual Values”*, publicada pela Fundação Cowles em 1950, e a prova da existência de equilíbrio num sistema de equilíbrio geral do tipo walrasiano em coautoria com Debreu, publicada em 1954.

Arrow e seus colaboradores publicaram três trabalhos importantes em relação ao Teorema de Kuhn-Tucker. Em dois artigos, *“Constraint Qualifications in Maximization Problems II”*³³ e *“Constraint Qualifications in Maximization Problems”*³⁴, analisaram algumas condições que, satisfeitas pelas restrições, implicam a validade da argumentação de Kuhn e Tucker via Lema de Farkas. As chamadas “qualificações das restrições” foram insuficientemente analisadas por Kuhn e Tucker³⁵.

No artigo *“Quasi-Concave Programming”*, Arrow mostrou que as condições de quase-concavidade e quase-convexidade tornam as

condições de Kuhn-Tucker suficientes para ótimos. No teorema 5 do artigo, apresenta pela primeira vez um critério suficiente para a quase-concavidade de uma função definida em $\mathbb{R}^n$, concluindo a análise iniciada por Hotelling em 1935³⁶ (ARROW, 1961a).

No artigo de 1927 sobre jogos de soma-zero, von Neumann apresentou duas condições que definem funções quase-côncavas e quase-convexas. Entretanto, foram formuladas como condições técnicas necessárias à argumentação e sem maior destaque. Na década de 1930, o matemático e economista italiano Bruno de Finetti (1906-1985) estudou funções quase-côncavas sob o ponto de vista da teoria econômica, inspirado pelo trabalho de Pareto em relação às curvas de indiferenças³⁷. As expressões “função quase-côncava” e “função quase-convexa” surgem na década de 1950. Elas aparecem no artigo do Nikaidô, que generaliza o Teorema Minimax³⁸, por exemplo. Nas notas “*Convex Cones, Sets and Functions*”, resultado dos seminários ministrados pelo matemático Werner Fenchel (1905-1988) em Princeton, em 1953, o autor define “função quase-convexa”³⁹.

¹ DEBREU, G. “Definite and Semidefinite Quadratic Forms”. **Econometrica**, v. 20, nº 2 (1952), pp. 295-300.

² MANN, H. “Quadratic forms with Linear Constraints”. **The American Mathematical Monthly**, v. 50, nº 7 (1943), pp. 430-433.

³ SYLVESTER, J. J. “A Demonstration of the Theorem that Every Homogeneous Quadratic Polynomial is Reducible by Real Orthogonal Substitutions to the Form of a Sum of Positive and Negative Squares”. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, Series 4, v. 4, nº 23 (1852), pp. 138-142.

⁴ COTTLE, R. “Manifestation of the Schur Complement”. **Linear Algebra and its Application**, v. 8 (1974), p. 190.

⁵ GANTMACHER, F. R. **Matrix Theory**, v. 1. New York: Chelsea Publishing Company, 1959, p. 302

- ⁶ DEBREU, G. "Definite and Semidefinite Quadratic Forms". **Econometrica**, v. 20, n° 2 (1952), pp. 295- 296.
- ⁷ VÄLIAHO, H. "On the Definity of Quadratic Forms Subject to Linear Constraints". **Journal of Optimization Theory and Applications**, v. 38 (1982), pp. 143-145.
- ⁸ GANTMACHER, F. R. **Matrix Theory**, v. 1. New York: Chelsea Publishing Company, 1959, p. 46.
- ⁹ COTTLE, R. "Manifestation of the Schur Complement". **Linear Algebra and its Application**, v. 8 (1974), p. 190.
- ¹⁰ ARROW, K. J.; ENTHOVEN, A. C. "Quasi-Concave Programming". **Econometrica**, v. 29, n° 4 (1961), pp. 779-800.
- ¹¹ HANCOCK, H. **Theory of Maxima and Minima**. EUA: Ginn Company, 1917.
- ¹² MOORE, C. Harris Hancock – In Memoriam. **Bulletin of the American Mathematical Society**, v. 50 (1944), pp. 812-815.
- ¹³ HANCOCK, H. **Theory of Maxima and Minima**. EUA: Ginn Company, 1917, p. iv.
- ¹⁴ MANN, H. "Quadratic forms with Linear Constraints". **The American Mathematical Monthly**, v. 50, n° 7 (1943), p. 430.
- ¹⁵ ARCHBOLD, J. W. **Algebra**. London: Sir Isaac Pitman & Sons, LTD., 1958, pp. 197-198.
- ¹⁶ HOTELLING, H. "Demand Functions with Limited Budgets". **Econometrica**, v. 3, n° 1 (1935), pp. 66-78.
- ¹⁷ SCHULTZ, H. **The Theory and Measurement of Demand**. Chicago: Chicago University Press, 1938.
- ¹⁸ LOUÇÃ, F. **The Years of High Econometrics: A Short History of the generation that Reinvented Economics**. NY: Routledge, 2007, p. 7.
- ¹⁹ MIROWSKI, P. **Machine Dreams: Economic Becomes a Cyborg Science**. NY: Cambridge University Press, 2002, pp. 202-203.
- ²⁰ SLUTSKY, E. "Sulla Teoria del Bilancio del Consumatore". **Giornale degli Economisti e Rivista di Statistica**, v. 51, n°1 (1915).
- ²¹ ALLEN, R. "Professor Slutsky's Theory of Consumers' Choice". **The Review of Economic Studies**, v. 3, n° 2 (1936).
- ²² HICKS, J. R. **Value and Capital: An Inquiry into some Fundamental Principles of Economic Theory**. Oxford: Oxford University Press, Second Edition, 1948.
- ²³ ARROW, K. J.; ENTHOVEN, A. C. "Quasi-Concave Programming". **Econometrica**, v. 29, n° 4 (1961).
- ²⁴ MANN, H. "Quadratic forms with Linear Constraints". **The American Mathematical Monthly**, v. 50, n° 7 (1943), pp. 430-433.
- ²⁵ NEWMAN, P. (ed). **Readings in Mathematical Economics**, v. 1. Baltimore: The Johns Hopkins Press, 1968, pp. 81-84.

- ²⁶ OLSON, J. Henry B. Mann in Hans Zassenhauss (ed.), **Number Theory and Algebra, Collected Papers Dedicated to Henry Mann, Arnold Ross and Olga Taussky-Todd**. NY: Academic Press, 1977.
- ²⁷ SAMUELSON, P. A. **Foundations of Economic Analysis**. New York: Atheneum, 1974.
- ²⁸ DEBREU, G. Definite and Semidefinite Quadratic Forms. **Econometrica**, v. 20, n. 2, 1952, p. 295-300.
- ²⁹ Ibidem
- ³⁰ Ibidem
- ³¹ Ibidem
- ³² SHILOV, G. **Linear Algebra**. New York: Dover Publications, INC., 1977.
- ³³ ARROW, K.; UZAWA, H. "Constraint Qualifications in Maximization Problems II". **Tecnical Report n. 84, Institute for Mathematical Studies in Social Sciences**, Stanford University, 1960.
- ³⁴ ARROW, K.; HURWICZ, L.; UZAWA, H. "Constraint Qualifications in Maximization Problems". **Naval Research Logistics**, v. 8, n° 2 (1961b).
- ³⁵ KUHN, H. "Nonlinear Programming: a Historical View". In: LEMKE, C. E.; COTTLE, R. W. (Eds.). **Nonlinear Programming: Proceedings of the SIAM-AMS Symposia**, v. 9, 1975.
- ³⁶ ARROW, K. J., ENTHOVEN, A. C. Quase-Concave Programming. **Econometrica**, v. 29, n. 4, 1961, p. 779-800.
- ³⁷ GUERRAGGIO, A.; MOLHO, E. "The Origins of Quasi-Concavity: a Development between Mathematics and Economics". **Historia Mathematica**, v. 31 (2004).
- ³⁸ NIKAIDÔ, H. "On von Neumann's Minimax Theorem". **Pacific Journal of Mathematics**, v. 4, n° 1 (1954).
- ³⁹ FENCHEL, W. **Convex Sets, Cones and Functions**. Princeton University, 1956.

Lição OITO

Envelopes e Estática Comparativa

Sumário

Lição 8 – Envelopes e Estática Comparativa

1. Introdução
 2. Teorema da função implícita
 3. Teoremas do envelope
 - 3.1 Teorema 2: Teorema do Envelope (caso sem restrições)
 - 3.1.1 Lema de Hotelling
 - 3.2 Teorema 3: Teorema do Envelope (caso com restrições)
 - 3.3 Envelopes, o teorema de Kuhn-Tucker e programação linear
 4. Estática comparativa
 - 4.1 A Metodologia primal-dual de Eugene Silberberg¹⁷ (caso sem restrições)
 - 4.3 Minimização de custos
 - 4.4 Teoria do consumidor: efeito-renda e efeito-substituição
 - 4.5 A curva de renda-consumo
 - 4.6 A curva de demanda-preço
 - 4.7 A metodologia primal-dual de Eugene Silberberg¹⁹ (caso com restrições)
 - 4.8 Explicando Silberberg
 5. Princípio de Le Chatelier-Samuelson
 - 5.1 Samuelson e um teorema “bem conhecido”
 - 5.2 Os artigos de Silberberg e o princípio de Le Chatelier-Samuelson
- Finalizando

Lição 8

Envelopes e Estática Comparativa

“...as universidades têm disciplinas, o mundo real tem problemas.”¹

1. Introdução

As lições de um a cinco visam estabelecer as condições de primeira ordem. Essas condições relacionam os valores dos ótimos das variáveis de escolha e os parâmetros exógenos. As condições de segunda ordem asseguram que as relações de primeira ordem são as relações corretas. Com esse objetivo, desenvolvemos os conceitos relacionados à convexidade e estabelecemos as condições de segunda ordem nas lições seis e sete.

Dada essa estrutura, o que podemos ainda obter é como as variáveis de escolha dependem dos parâmetros exógenos. Esse processo é chamado de “Estática Comparativa” e é um dos objetivos desta lição.

O outro objetivo trata dos chamados “Teoremas do Envelope”, que complementam a Estática Comparativa estabelecendo regras para obter as derivadas da chamada “Função Valor” em relação aos parâmetros exógenos, isto é, o valor ótimo da função objetivo.

Para desenvolvê-los de maneira adequada, precisaremos de uma breve exposição do chamado “Teorema da Função Implícita” que, apesar de já termos utilizado em tópicos anteriores, achamos melhor reservar para esta lição.

Por fim, aplicaremos os resultados desta lição para complementar a dualidade em problemas lineares em relação ao Esquema Simplex.

A Estática Comparativa está muito associada ao nome de Paul Samuelson (1915-2009) por causa do livro “*Economic Analysis*”², mas Samuelson não é o seu criador. Muitos economistas, entre os mais conhecidos estão Harold Hotelling e o ucraniano Eugen Slutsky (1880-1948), desenvolveram a Estática Comparativa.

Samuelson tem o mérito de ter explicitado que as relações que constituem a Estática Comparativa derivam, primeiramente, dos sistemas de equações que resultam das condições de primeira ordem de problemas de otimização e, em segundo lugar, das condições de segunda ordem que esses problemas implicam. Isso estava latente, por exemplo, na colaboração entre Schultz e Hotelling, como comentamos na lição anterior. Em outras palavras, todas as expressões da Estática Comparativa são obtidas das relações estabelecidas entre as variáveis através dos pontos críticos das funções envolvidas ou das definições das formas quadráticas subentendidas pela formulação do problema. Esse é o significado da frase inicial da Introdução do “Análise”: “A existência de analogias entre as características centrais de várias teorias implica a existência de uma teoria geral que subjaz às teorias particulares e as unifica com relação a essas características centrais”³.

A “Metodologia Primal-Dual”⁴, de Eugene Silberberg (1940), considera um problema de otimização em substituição ao problema original, no qual as variáveis endógenas e exógenas são tratadas como variáveis endógenas nesse novo problema. Como o problema exige que os pontos críticos sejam mínimos ou máximos, isso implica que a derivada segunda do lagrangeano deve satisfazer as condições adequadas quanto à definição. Todas as relações possíveis entre as derivadas envolvidas, tanto das soluções quanto da função-valor, resultam da definição necessária da derivada segunda do lagrangeano. Corretamente, do ponto de vista metodológico, Silberberg analisa as consequências matemáticas que resultam da formulação do problema, ao passo que a postulação tradicional impõe, *a priori*, ao modelo

determinadas condições matemáticas objetivando que os problemas propostos pelo modelo tenham as soluções esperadas.

A partir disso, o leitor compreende o porquê do empenho anterior neste livro em estabelecer tão cuidadosamente as condições de primeira e segunda ordem. Uma vez que em Teoria Econômica não existem formas funcionais definidas que determinem as variáveis, temos que nos satisfazer com as propriedades gerais que essas variáveis exibem em conjunto – por exemplo, que certos determinantes são não-positivos ou não-negativos – e, a partir disso, relacioná-las. Esse também é o limite possível para este tipo de teoria baseada em princípios muito gerais e em otimização.

2. Teorema da função implícita

Nós estamos lidando com o Teorema da Função Implícita quando “resolvemos” um sistema linear por linha-redução expressando algumas variáveis, ditas dependentes, do sistema em função das demais, ditas independentes. Observe que “resolver” o sistema significa que podemos atribuir valores arbitrários às variáveis independentes e calcular, através das equações obtidas pela linha-redução, os valores das variáveis dependentes.

Exemplo 1: Seja o sistema

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= 3 \cdot x_1 - 5 \cdot x_2 + 2 \cdot x_3 + 4 \cdot x_4 = 0, \\ f_2(x_1, x_2, x_3, x_4) &= 6 \cdot x_1 - 10 \cdot x_2 - 4 \cdot x_3 - 6 \cdot x_4 = 0. \end{aligned}$$

Por linha-redução obtemos

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{pmatrix} = \begin{pmatrix} 3 & -5 & 2 & 4 \\ 6 & -10 & -4 & -6 \end{pmatrix} \sim \begin{pmatrix} 1 & -5/3 & 0 & 1/2 \\ 0 & 0 & 1 & 7/4 \end{pmatrix},$$

portanto, a solução é dada por

$$\begin{aligned} x_1 &= \frac{5}{3} \cdot x_2 - \frac{1}{2} \cdot x_4 \\ x_3 &= -\frac{7}{4} \cdot x_4. \end{aligned}$$

Observe que fomos automaticamente conduzidos a expressar x_1 e x_3 como funções de x_2 e x_4 , e não x_1 e x_2 como funções de x_3 e x_4 ! É fácil ver porque isso ocorreu. Se pudéssemos expressar x_1 e x_2 como funções de x_3 e x_4 , então a matriz linha-reduzida seria da forma $\begin{pmatrix} 1 & 0 & b_{13} & b_{14} \\ 0 & 1 & b_{23} & b_{24} \end{pmatrix}$, o que implica que a submatriz $\begin{pmatrix} 3 & -5 \\ 6 & -10 \end{pmatrix}$ seria linha-equivalente à matriz $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ e, portanto, inversível. Entretanto, essa submatriz não é inversível, pois $\begin{vmatrix} 3 & -5 \\ 6 & -10 \end{vmatrix} = 0$. A linha-redução se dá sobre a submatriz $\begin{pmatrix} 3 & 2 \\ 6 & -4 \end{pmatrix}$, porque $\begin{vmatrix} 3 & 2 \\ 6 & -4 \end{vmatrix} \neq 0$. Facilmente percebemos que as variáveis x_i e x_j podem ser expressas através das demais se $\begin{vmatrix} a_{1i} & a_{1j} \\ a_{2i} & a_{2j} \end{vmatrix} \neq 0$. Por exemplo, $\begin{vmatrix} 2 & 4 \\ -4 & -6 \end{vmatrix} = 4$, então podemos pôr as variáveis x_3 e x_4 como funções das variáveis x_1 e x_2 :

$\begin{pmatrix} 3 & -5 & 2 & 4 \\ 6 & -10 & -4 & -6 \end{pmatrix} \rightarrow \begin{pmatrix} -21/2 & 35/2 & 1 & 0 \\ 6 & -10 & 0 & 1 \end{pmatrix}$, o que resulta em:

$$\begin{aligned} x_3 &= \frac{21}{2} \cdot x_1 - \frac{35}{2} \cdot x_2 \\ x_4 &= -6 \cdot x_1 + 10 \cdot x_2. \end{aligned}$$

Observe que

$$\begin{aligned} 3 \cdot x_1 - 5 \cdot x_2 + 2 \cdot \left(\frac{21}{2} \cdot x_1 - \frac{35}{2} \cdot x_2 \right) + 4 \cdot (-6 \cdot x_1 + 10 \cdot x_2) &= 0 \\ 6 \cdot x_1 - 10 \cdot x_2 - 4 \cdot \left(\frac{21}{2} \cdot x_1 - \frac{35}{2} \cdot x_2 \right) - 6 \cdot (-6 \cdot x_1 + 10 \cdot x_2) &= 0. \end{aligned}$$

Ou seja, expressamos implicitamente as variáveis x_3 e x_4 como funções das variáveis x_1 e x_2 nas equações iniciais. A condição que nos permite esse

procedimento é $\begin{vmatrix} 2 & 4 \\ -4 & -6 \end{vmatrix} = \begin{vmatrix} \frac{\partial f_1}{\partial x_3} & \frac{\partial f_1}{\partial x_4} \\ \frac{\partial f_2}{\partial x_3} & \frac{\partial f_2}{\partial x_4} \end{vmatrix} \neq 0$. Mais ainda. Observe que $\frac{\partial x_3}{\partial x_1} = \frac{21}{2}$, $\frac{\partial x_3}{\partial x_2} = -\frac{35}{2}$, $\frac{\partial x_4}{\partial x_1} = -6$ e

$\frac{\partial x_4}{\partial x_2} = 10$, ou seja,

$$\begin{pmatrix} \frac{\partial x_3}{\partial x_1} & \frac{\partial x_3}{\partial x_2} \\ \frac{\partial x_4}{\partial x_1} & \frac{\partial x_4}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \frac{21}{2} & -\frac{35}{2} \\ -6 & 10 \end{pmatrix} = - \begin{pmatrix} 2 & 4 \\ -4 & -6 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 3 & -5 \\ 6 & -10 \end{pmatrix} = - \begin{pmatrix} -\frac{3}{2} & -1 \\ 1 & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 3 & -5 \\ 6 & -10 \end{pmatrix}.$$

Portanto, observemos que

$$\begin{pmatrix} \frac{\partial x_3}{\partial x_1} & \frac{\partial x_3}{\partial x_2} \\ \frac{\partial x_1}{\partial x_3} & \frac{\partial x_2}{\partial x_3} \\ \frac{\partial x_4}{\partial x_1} & \frac{\partial x_4}{\partial x_2} \end{pmatrix} = - \begin{pmatrix} \frac{\partial f_1}{\partial x_3} & \frac{\partial f_1}{\partial x_4} \\ \frac{\partial f_2}{\partial x_3} & \frac{\partial f_2}{\partial x_4} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}.$$

Suponhamos que para qualquer $(x, y) = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$ que satisfaça o conjunto de equações abaixo

$$\begin{aligned} f_1(x_1, \dots, x_n, y_1, \dots, y_m) &= c_1, \\ f_2(x_1, \dots, x_n, y_1, \dots, y_m) &= c_2, \\ &\vdots \\ f_m(x_1, \dots, x_n, y_1, \dots, y_m) &= c_m, \end{aligned}$$

tenhamos

$$\begin{vmatrix} \frac{\partial f_1}{\partial y_1}(x, y) & \frac{\partial f_1}{\partial y_2}(x, y) & \cdots & \frac{\partial f_1}{\partial y_m}(x, y) \\ \frac{\partial f_2}{\partial y_1}(x, y) & \frac{\partial f_2}{\partial y_2}(x, y) & \cdots & \frac{\partial f_2}{\partial y_m}(x, y) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial y_1}(x, y) & \frac{\partial f_m}{\partial y_2}(x, y) & \cdots & \frac{\partial f_m}{\partial y_m}(x, y) \end{vmatrix} \neq 0.$$

Observe que a condição acima é a condição necessária para pôr as variáveis $y_1, \dots, y_m$ como funções das variáveis $x_1, \dots, x_n$, no caso em que as funções f_i sejam lineares. Se for possível pôr $y_1, \dots, y_m$ como funções das variáveis $x_1, \dots, x_n$, obtemos

$$\begin{aligned} f_1(x_1, \dots, x_n, y_1(x_1, \dots, x_n), \dots, y_m(x_1, \dots, x_n)) &= c_1, \\ f_2(x_1, \dots, x_n, y_1(x_1, \dots, x_n), \dots, y_m(x_1, \dots, x_n)) &= c_2, \\ &\vdots \\ f_m(x_1, \dots, x_n, y_1(x_1, \dots, x_n), \dots, y_m(x_1, \dots, x_n)) &= c_m. \end{aligned}$$

Pela regra da cadeia, obtemos para $j = 1, \dots, n$

$$\begin{aligned} \frac{\partial f_1}{\partial x_j}(x, y(x)) + \frac{\partial f_1}{\partial y_1}(x, y(x)) \cdot \frac{\partial y_1}{\partial x_j}(x) + \cdots + \frac{\partial f_1}{\partial y_m}(x, y(x)) \cdot \frac{\partial y_m}{\partial x_j}(x) &= 0, \\ \frac{\partial f_2}{\partial x_j}(x, y(x)) + \frac{\partial f_2}{\partial y_1}(x, y(x)) \cdot \frac{\partial y_1}{\partial x_j}(x) + \cdots + \frac{\partial f_2}{\partial y_m}(x, y(x)) \cdot \frac{\partial y_m}{\partial x_j}(x) &= 0, \\ &\vdots \\ \frac{\partial f_m}{\partial x_j}(x, y(x)) + \frac{\partial f_m}{\partial y_1}(x, y(x)) \cdot \frac{\partial y_1}{\partial x_j}(x) + \cdots + \frac{\partial f_m}{\partial y_m}(x, y(x)) \cdot \frac{\partial y_m}{\partial x_j}(x) &= 0. \end{aligned}$$

Observe que as equações acima podem ser condensadas na expressão matricial abaixo:

$$\underbrace{\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}}_{m \times n} + \underbrace{\begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} \end{pmatrix}}_{m \times m} \cdot \underbrace{\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}}_{m \times n} = \underbrace{\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}}_{m \times n}$$

Portanto,
$$\begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} = - \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$
 Como, por hipótese, $\begin{vmatrix} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} \end{vmatrix} \neq 0,$

então

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_1}(x) & \dots & \frac{\partial y_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1}(x) & \dots & \frac{\partial y_m}{\partial x_n}(x) \end{pmatrix} = - \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(x, y) & \dots & \frac{\partial f_1}{\partial y_m}(x, y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(x, y) & \dots & \frac{\partial f_m}{\partial y_m}(x, y) \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x, y) & \dots & \frac{\partial f_1}{\partial x_n}(x, y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x, y) & \dots & \frac{\partial f_m}{\partial x_n}(x, y) \end{pmatrix},$$

que é a expressão que obtivemos no exemplo anterior. Vamos empregar uma notação mais condensada:

$$\partial_x y = \begin{pmatrix} \frac{\partial y_1}{\partial x_1}(x) & \dots & \frac{\partial y_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1}(x) & \dots & \frac{\partial y_m}{\partial x_n}(x) \end{pmatrix}, \partial_y f = \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(x, y) & \dots & \frac{\partial f_1}{\partial y_m}(x, y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(x, y) & \dots & \frac{\partial f_m}{\partial y_m}(x, y) \end{pmatrix}, \partial_x f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x, y) & \dots & \frac{\partial f_1}{\partial x_n}(x, y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x, y) & \dots & \frac{\partial f_m}{\partial x_n}(x, y) \end{pmatrix}, \text{ ou seja, } \partial_x y = -(\partial_y f)^{-1} \cdot (\partial_x f).$$

Para a validade do resultado abaixo, não é necessário que as funções $f_1, f_2, \dots, f_m$ sejam lineares.

Teorema 1 - Teorema da Função Implícita

Seja $f: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$ continuamente diferenciável, tal que $f(x^*, y^*) = c$ e $|\partial_y f(x^*, y^*)| \neq 0$, então existem vizinhanças abertas $V(x^*)$ e $V(y^*)$ e uma função $y: V(x^*) \rightarrow V(y^*)$ continuamente diferenciável, tal que $y(x^*) = y^*$, $f(x, y(x)) = c$ e $\partial_x y(x) = -(\partial_y f(x, y(x)))^{-1} \cdot (\partial_x f(x, y(x)))$, para $x \in V(x^*)$.

Na situação geral em que não estamos supondo as funções $f_1, f_2, \dots, f_m$ lineares, como no exemplo anterior, não podemos assegurar que as funções $y_1, y_2, \dots, y_m$ estejam definidas em todo o espaço $\mathbb{R}^m$, mas numa vizinhança de qualquer (x^*, y^*) , tal que $f(x^*, y^*) = c$.

Em 1965, David Gale (1921-2008) e Hukukane Nikaidô (1923-2001) estabeleceram certas condições que garantem a existência de funções

implícitas em regiões com características particulares⁵.

Em geral, o conjunto de variáveis dependentes $y_1, y_2, \dots, y_m$ não é distinto, em termos notacionais, das variáveis independentes $x_1, \dots, x_n$. Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m < n$, fixado um $x^* = (x_1^*, x_2^*, \dots, x_n^*)$, digamos que exista uma submatriz de ordem m ,

$$\begin{vmatrix} \frac{\partial f_1}{\partial x_{i_1}}(x^*) & \frac{\partial f_1}{\partial x_{i_2}}(x^*) & \dots & \frac{\partial f_1}{\partial x_{i_m}}(x^*) \\ \frac{\partial f_2}{\partial x_{i_1}}(x^*) & \frac{\partial f_2}{\partial x_{i_2}}(x^*) & \dots & \frac{\partial f_2}{\partial x_{i_m}}(x^*) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_{i_1}}(x^*) & \frac{\partial f_m}{\partial x_{i_2}}(x^*) & \dots & \frac{\partial f_m}{\partial x_{i_m}}(x^*) \end{vmatrix} \neq 0.$$

Observe que estamos particionando o conjunto $\{1, 2, \dots, n\}$ em dois subconjuntos $\{i_1, i_2, \dots, i_m\}$ e $\{j_1, j_2, \dots, j_{n-m}\}$. O teorema da função implícita assegura que, nesse caso, na proximidade do ponto $x^* = (x_1^*, x_2^*, \dots, x_n^*)$, as variáveis $x_{i_1}, \dots, x_{i_m}$ podem ser postas em função das variáveis $x_{j_1}, \dots, x_{j_{n-m}}$ satisfazendo as equações

$$\begin{aligned} f_1(x_1, \dots, x_n) &= c_1, \\ f_2(x_1, \dots, x_n) &= c_2, \\ &\vdots \\ f_m(x_1, \dots, x_n) &= c_m. \end{aligned}$$

Ou seja, o conjunto dos pontos $x = (x_1, x_2, \dots, x_n)$, descritos pelas equações acima, é tal que, na proximidade de cada um destes pontos, $n-m$ variáveis independentes descrevem m variáveis dependentes. Isso define o que chamamos uma **superfície de dimensão $n-m$** no $\mathbb{R}^n$. Essa situação é a que encontramos no exemplo anterior, pois as equações

$$\begin{aligned} x_3 &= \frac{21}{2} \cdot x_1 - \frac{35}{2} \cdot x_2, \\ x_4 &= -6 \cdot x_1 + 10 \cdot x_2, \end{aligned}$$

definem um subespaço de dimensão 2 no $\mathbb{R}^4$ que satisfaz a nossa definição de superfície.

Exemplo 2: Num sistema linear com k equações e n incógnitas, cujo posto da matriz dos coeficientes é k , o processo de linha-redução pode ser feito através das variáveis que determinam uma submatriz de ordem k com determinante não nulo. Por exemplo, consideremos o sistema:

$$\begin{aligned} 3 \cdot x_1 - x_2 - 2 \cdot x_3 + 3 \cdot x_4 + 4 \cdot x_5 &= 1, \\ x_1 + 5 \cdot x_2 + 2 \cdot x_3 - x_4 + 7 \cdot x_5 &= 1. \end{aligned}$$

$$\text{posto} \begin{pmatrix} 3 & -1 & -2 & 3 & 4 \\ 1 & 5 & 2 & -1 & 7 \end{pmatrix} = 2$$

Como $\begin{vmatrix} -1 & 3 \\ 5 & -1 \end{vmatrix} \neq 0$, o processo de linha-redução pode ser feito sobre as variáveis x_2 e x_4 , resultando

$$\begin{aligned} x_2(x_1, x_3, x_5) &= \frac{2}{7} - \frac{3}{7} \cdot x_1 - \frac{2}{7} \cdot x_3 - \frac{25}{14} \cdot x_5, \\ x_4(x_1, x_3, x_5) &= \frac{3}{7} - \frac{8}{7} \cdot x_1 + \frac{4}{7} \cdot x_3 - \frac{27}{14} \cdot x_5. \end{aligned}$$

Observe que

$$\begin{pmatrix} \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_3} & \frac{\partial x_2}{\partial x_5} \\ \frac{\partial x_4}{\partial x_1} & \frac{\partial x_4}{\partial x_3} & \frac{\partial x_4}{\partial x_5} \end{pmatrix} = \begin{pmatrix} -\frac{3}{7} & -\frac{2}{7} & -\frac{25}{14} \\ -\frac{8}{7} & \frac{4}{7} & -\frac{27}{14} \end{pmatrix} = - \begin{pmatrix} -1 & 3 \\ 5 & -1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 3 & -2 & 4 \\ 1 & 2 & 7 \end{pmatrix},$$

exemplificando que $\partial_{x_1, x_3, x_5} x_2, x_4 = -(\partial_{x_2, x_4} f)^{-1} \cdot (\partial_{x_1, x_3, x_5} f)$.

Também podemos obter as derivadas acima derivando os sistemas $3 \cdot x_1 - x_2 - 2 \cdot x_3 + 3 \cdot x_4 + 4 \cdot x_5 = 1$, $x_1 + 5 \cdot x_2 + 2 \cdot x_3 - x_4 + 7 \cdot x_5 = 1$. Por exemplo, derivando em relação ao x_3 ,

obtemos $\begin{pmatrix} -1 & 3 \\ 5 & -1 \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial x_2}{\partial x_3} \\ \frac{\partial x_4}{\partial x_3} \end{pmatrix} = - \begin{pmatrix} -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$, ou seja, $\begin{pmatrix} \frac{\partial x_2}{\partial x_3} \\ \frac{\partial x_4}{\partial x_3} \end{pmatrix} = - \begin{pmatrix} -1 & 3 \\ 5 & -1 \end{pmatrix}^{-1} \begin{pmatrix} -2 \\ 2 \end{pmatrix}$. Ou ainda, por Cramer,

$\frac{\partial x_2}{\partial x_3} = \frac{\begin{vmatrix} 2 & 3 \\ -2 & -1 \end{vmatrix}}{\begin{vmatrix} -1 & 3 \\ 5 & -1 \end{vmatrix}} = -\frac{2}{7}$ e $\frac{\partial x_4}{\partial x_3} = \frac{\begin{vmatrix} -1 & 2 \\ 5 & -2 \end{vmatrix}}{\begin{vmatrix} -1 & 3 \\ 5 & -1 \end{vmatrix}} = \frac{4}{7}$. As outras duas colunas de $\begin{pmatrix} \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_3} & \frac{\partial x_2}{\partial x_5} \\ \frac{\partial x_4}{\partial x_1} & \frac{\partial x_4}{\partial x_3} & \frac{\partial x_4}{\partial x_5} \end{pmatrix}$ são obtidas de

maneira similar.

Exemplo 3: Num sistema linear com k equações e n incógnitas, cujo posto da matriz dos coeficientes é k , o processo de linha-redução pode ser feito através das variáveis que determinam uma submatriz de ordem k com determinante não nulo. Por exemplo, consideremos o sistema

$$\begin{aligned} f_1(x_1, x_2, x_3) &= 3 \cdot x_1 - x_2 - 2 \cdot x_3 + 3 \cdot x_4 + 4 \cdot x_5 = 1, \\ f_2(x_1, x_2, x_3) &= x_1 + 5 \cdot x_2 + 2 \cdot x_3 - x_4 + 7 \cdot x_5 = 1. \end{aligned}$$

Observe que o $\text{posto} \begin{pmatrix} 3 & -1 & -2 & 3 & 4 \\ 1 & 5 & 2 & -1 & 7 \end{pmatrix} = 2$, pois $\begin{vmatrix} 3 & -1 \\ 1 & 5 \end{vmatrix} \neq 0$, por exemplo. O processo de linha-redução pode ser feito sobre as variáveis x_1 e x_2 :

$$\left(\begin{array}{cccc|c} 3 & -1 & -2 & 3 & 4 & 1 \\ 1 & 5 & 2 & -1 & 7 & 1 \end{array} \right) \sim \left(\begin{array}{cccc|c} 1 & 0 & -\frac{1}{2} & \frac{7}{8} & \frac{27}{16} & \frac{3}{8} \\ 0 & 1 & \frac{1}{2} & -\frac{3}{8} & \frac{17}{16} & \frac{1}{8} \end{array} \right).$$

Portanto,

$$\begin{aligned} x_1(x_3, x_4, x_5) &= \frac{3}{8} + \frac{1}{2} \cdot x_3 - \frac{7}{8} \cdot x_4 - \frac{27}{16} \cdot x_5, \\ x_2(x_3, x_4, x_5) &= \frac{1}{8} - \frac{1}{2} \cdot x_3 + \frac{3}{8} \cdot x_4 - \frac{17}{16} \cdot x_5. \end{aligned}$$

Observe que
$$\begin{pmatrix} \frac{\partial x_1}{\partial x_3} & \frac{\partial x_1}{\partial x_4} & \frac{\partial x_1}{\partial x_5} \\ \frac{\partial x_2}{\partial x_3} & \frac{\partial x_2}{\partial x_4} & \frac{\partial x_2}{\partial x_5} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{7}{8} & -\frac{27}{16} \\ -\frac{1}{2} & \frac{3}{8} & -\frac{17}{16} \end{pmatrix} = - \begin{pmatrix} 3 & -1 \\ 1 & 5 \end{pmatrix}^{-1} \cdot \begin{pmatrix} -2 & 3 & 4 \\ 2 & -1 & 7 \end{pmatrix}.$$

O processo de linha-redução pode ser feito sobre as variáveis x_2 e x_4 , pois $\begin{vmatrix} -1 & 3 \\ 5 & -1 \end{vmatrix} \neq 0$ resultando

$$\begin{aligned} x_2(x_1, x_3, x_5) &= \frac{2}{7} - \frac{3}{7} \cdot x_1 - \frac{2}{7} \cdot x_3 - \frac{25}{14} \cdot x_5, \\ x_4(x_1, x_3, x_5) &= \frac{3}{7} - \frac{8}{7} \cdot x_1 + \frac{4}{7} \cdot x_3 - \frac{27}{14} \cdot x_5. \end{aligned}$$

Observe que
$$\begin{pmatrix} \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_3} & \frac{\partial x_2}{\partial x_5} \\ \frac{\partial x_4}{\partial x_1} & \frac{\partial x_4}{\partial x_3} & \frac{\partial x_4}{\partial x_5} \end{pmatrix} = \begin{pmatrix} -\frac{3}{7} & -\frac{2}{7} & -\frac{25}{14} \\ -\frac{8}{7} & \frac{4}{7} & -\frac{27}{14} \end{pmatrix} = - \begin{pmatrix} -1 & 3 \\ 5 & -1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 3 & -2 & 4 \\ 1 & 2 & 7 \end{pmatrix}.$$

Adotemos a seguinte notação: $\partial_{x_1, x_3, x_5} f = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_3} & \frac{\partial f}{\partial x_5} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_3} & \frac{\partial f_2}{\partial x_5} \end{pmatrix}, \quad \partial_{x_2, x_4} f = \begin{pmatrix} \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial x_4} \\ \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_4} \end{pmatrix} \quad \text{e}$

$\partial_{x_1, x_3, x_5} x_2, x_4 = \begin{pmatrix} \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_3} & \frac{\partial x_2}{\partial x_5} \\ \frac{\partial x_4}{\partial x_1} & \frac{\partial x_4}{\partial x_3} & \frac{\partial x_4}{\partial x_5} \end{pmatrix}$, então $\partial_{x_1, x_3, x_5} x_2, x_4 = -(\partial_{x_2, x_4} f)^{-1} \cdot (\partial_{x_1, x_3, x_5} f)$. Similarmente, no primeiro caso,

$$\partial_{x_3, x_4, x_5} x_1, x_2 = -(\partial_{x_1, x_2} f)^{-1} \cdot (\partial_{x_3, x_4, x_5} f).$$

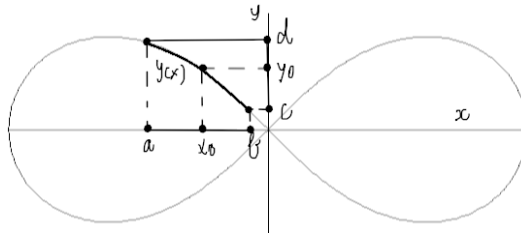
Exemplo 4: Consideremos o ponto $(x, y) \in \mathbb{R}^2$. A distância de (x, y) ao ponto $(-a, 0)$ é $r_1 = \sqrt{(x+a)^2 + y^2}$ e a distância ao $(a, 0)$ é $r_2 = \sqrt{(x-a)^2 + y^2}$. A curva, chamada lemniscata, definida pelos pontos $(x, y) \in \mathbb{R}^2$, tais que $r_1 \cdot r_2 = a^2$. Portanto, $r_1^2 \cdot r_2^2 = a^4$, ou seja, $((x+a)^2 + y^2) \cdot ((x-a)^2 + y^2) = a^4$, ou, $(x^2 + y^2)^2 - 2 \cdot a^2 \cdot (x^2 - y^2) = 0$. Abaixo desenhamos a lemniscata tal que $a=3$. Essa curva é definida pela isoquanta de nível zero da função $f(x, y) = x^4 + 2 \cdot x^2 \cdot y^2 - 18 \cdot x^2 + y^4 + 18 \cdot y^2 = 0$. $\frac{\partial f}{\partial y}(x, y) = 4 \cdot y \cdot (x^2 + y^2 + 9) \neq 0$, exceto nos pontos em que $y=0$, que são $(-3 \cdot \sqrt{2}, 0)$, $(0, 0)$ e $(3 \cdot \sqrt{2}, 0)$. Portanto, se (x_0, y_0) , $y_0 \neq 0$, existem intervalos (a, b) , (c, d) e uma função $y(x)$, tal que $y: (a, b) \rightarrow (c, d)$ e $y(x_0) = y_0$ $f(x, y(x)) = x^4 + 2 \cdot x^2 \cdot (y(x))^2 - 18 \cdot x^2 + (y(x))^4 + 18 \cdot (y(x))^2 = 0$, ou seja, $y(x)$ está implicitamente definida na equação $f(x, y(x)) = 0$. A derivada de $y'(x)$

é obtida derivando $f(x, y(x)) = 0$: $\frac{\partial f}{\partial x}(x, y(x)) + \frac{\partial f}{\partial y}(x, y(x)) \cdot y'(x) = 0$, pela regra da cadeia, ou

$$y'(x) = -\frac{\frac{\partial f}{\partial x}(x, y(x))}{\frac{\partial f}{\partial y}(x, y(x))} = -\frac{x \cdot (x^2 + y^2 - 9)}{y \cdot (x^2 + y^2 + 9)},$$

que é a expressão obtida anteriormente.

$\partial_x y(x) = -(\partial_y f(x, y(x)))^{-1} \cdot (\partial_x f(x, y(x)))$. Veja o desenho:



Um resultado que empregamos anteriormente e que mostramos via o Teorema da Função Implícita é aquele que assegura que se um vetor qualquer $v \in \mathbb{R}^n$, tal que $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, existe uma trajetória $\varphi: (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$, para algum $\varepsilon > 0$, tal que $\varphi(0) = x^*, g_1(\varphi(t)) = 0, g_2(\varphi(t)) = 0, \dots, g_m(\varphi(t)) = 0$ e $\varphi'(0) = v$. Em outras palavras, estamos mostrando que qualquer vetor ortogonal aos gradientes $\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_m(x^*)$ é vetor tangente à superfície definida por $g_1(x) = 0, g_2(x) = 0, \dots, g_m(x) = 0$ em x^* .

Como estamos supondo que o sistema de equações $g_1(x) = 0, g_2(x) = 0, \dots, g_m(x) = 0$ define uma superfície, então posto $\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(x^*) & \dots & \frac{\partial g_m}{\partial x_n}(x^*) \end{pmatrix} = m$. Isso significa que na vizinhança de x^* podemos expressar, digamos, as variáveis $x_1, x_2, \dots, x_m$ como funções das demais $x_{m+1}, x_{m+2}, \dots, x_n$ nas equações $g_1(x) = 0, g_2(x) = 0, \dots, g_m(x) = 0$.

Para $i = m+1, \dots, n$, considere as trajetórias sobre a superfície: $\varphi_i(t) = (x_1(x_{m+1}^*, \dots, x_{m+i}^* + t, \dots, x_n^*), \dots, x_m(x_{m+1}^*, \dots, x_{m+i}^* + t, \dots, x_n^*), x_{m+1}^*, \dots, x_{m+i}^* + t, \dots, x_n^*)$, então $\varphi_i(0) = x^*$ e $\varphi_i'(0) = \left(\frac{\partial x_1}{\partial x_{m+i}}(x_{m+1}^*, \dots, x_n^*), \dots, \frac{\partial x_m}{\partial x_{m+i}}(x_{m+1}^*, \dots, x_n^*), 0, \dots, 0, 1, 0, \dots, 0 \right)$. Os vetores $\varphi_1'(0), \varphi_2'(0), \dots, \varphi_{n-m}'(0)$ são linearmente independentes e como a dimensão do espaço dos vetores ortogonais aos vetores $\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_m(x^*)$ tem dimensão $n-m$, então constituem uma base de vetores tangentes para esse espaço.

Seja $v \in \mathbb{R}^n$, tal que $\nabla g_1(x^*) \cdot v = 0, \nabla g_2(x^*) \cdot v = 0, \dots, \nabla g_m(x^*) \cdot v = 0$, então existem $\lambda_1, \dots, \lambda_{n-m} \in \mathbb{R}$, tais que $v = \lambda_1 \cdot \varphi_1'(0) + \dots + \lambda_{n-m} \cdot \varphi_{n-m}'(0)$. Seja a trajetória $\varphi(t) =$

$$(x_1^*(x_{m+1}^* + \lambda_1 \cdot t, \dots, x_n^* + \lambda_n \cdot t), \dots, x_m^*(x_{m+1}^* + \lambda_1 \cdot t, \dots, x_n^* + \lambda_n \cdot t), x_{m+1}^* + \lambda_1 \cdot t, \dots, x_n^* + \lambda_n \cdot t), \text{então } \varphi'(0) = v = \lambda_1 \cdot \varphi_1'(0) + \dots + \lambda_{n-m} \cdot \varphi_{n-m}'(0).$$

3. Teoremas do envelope

No século 20, a história desses teoremas começa em 1947 com o célebre livro “*Foundations of Economic Analysis*”⁶ (1949), de Paul Samuelson (1915-2009): “para que a análise seja útil ela tem que fornecer informações a respeito do modo em que nossas quantidades de equilíbrio irão variar como resultado das variações dos parâmetros tomados como dados independentes”⁷.

No capítulo 3, intitulado “A Teoria do Comportamento Maximizante”, parágrafo “O Deslocamento da Quantidade Maximizada”, Samuelson aborda a questão das derivadas em relação a um parâmetro exógeno α da função $z(\alpha)$ obtida pelo processo de maximização nas variáveis de escolha $x_1, \dots, x_n$ de uma função de $f(x_1, \dots, x_n, \alpha)$. O parâmetro exógeno α é mantido fixo no processo de maximização. Se $x^*(\alpha) = (x_1^*(\alpha), \dots, x_n^*(\alpha))$ é um mínimo ou máximo local de $z(\alpha) = \max_{x_1, \dots, x_n} f(x_1, \dots, x_n, \alpha)$, então $\frac{\partial z}{\partial \alpha}(\alpha) = \frac{\partial f}{\partial \alpha}(x^*(\alpha), \alpha)$. Após justificar a expressão anterior, Samuelson comenta que se obtém uma expressão similar para o caso da otimização restrita. Ou seja, para o caso:

$$\begin{aligned} \min/\max_{x_1, \dots, x_n} \quad & f(x_1, \dots, x_n, \alpha) = z(\alpha) \\ \text{sa} \quad & g_1(x_1, \dots, x_n, \alpha) \leq 0; \\ & \vdots \\ & g_k(x_1, \dots, x_n, \alpha) \leq 0; \\ & h_1(x_1, \dots, x_n, \alpha) \geq 0; \\ & \vdots \\ & h_l(x_1, \dots, x_n, \alpha) \geq 0. \end{aligned}$$

Obviamente, podemos ter um vetor de parâmetros exógenos $\alpha_1, \dots, \alpha_m$.

Sobre a razão do nome “Teoremas do Envelope”, Samuelson explica: “esta é a relação familiar de tangência entre o envelope de uma família de curvas e as curvas que o tocam”⁸.

O problema básico dos teoremas do envelope é dar conta de como os parâmetros exógenos impactam a função valor z . Por essa razão, alguns

autores referem-se a esses resultados como “Análise de Sensibilidade”. Richard Cornes, no livro “*Duality and Modern Economics*”⁹ (1992), nomeia os teoremas acima de “Teoremas do Envelope de 1ª Ordem”.

Relacionado ao resultado anterior, Samuelson trata outro problema mais complicado, a que chamou “Princípio de Le Chatelier” e que, atualmente, nos referimos como “Princípio de Le Chatelier-Samuelson”. Digamos que algumas variáveis de escolha estejam fixadas nos seus valores ótimos $x_1^*(\alpha), \dots, x_k^*(\alpha)$, e consideremos a função $\hat{z}(x_1, \dots, x_n, \alpha) = \min_{x_{k+1}, \dots, x_n} \max_{x_1, \dots, x_k} f(x_1, \dots, x_n, \alpha)$, as variáveis de escolha são as variáveis $x_i \neq x_i^*(\alpha), \dots, x_k(\alpha)$. A função $\hat{z}$ chamamos de “curto prazo”, enfatizando que o conjunto das variáveis de escolha foi restringido em relação ao conjunto das variáveis de escolha da função $z(\alpha) = \min_{x \in (x_1, \dots, x_n)} \max f(x_1, \dots, x_n, \alpha)$, dita de “longo prazo”. O Princípio de Le Chatelier-Samuelson analisa a relação entre as derivadas $\frac{\partial z}{\partial \alpha}$ e $\frac{\partial \hat{z}}{\partial \alpha}$.

Um tanto vagamente, esse problema se relaciona a um princípio formulado, em 1884, pelo químico Henri Louis Le Châtelier (1850-1936) e que leva seu nome¹⁰. Richard Cornes chama esses teoremas de “Teoremas do Envelope de 2ª ordem”. Mais geralmente, a questão acima trata do impacto sobre os valores ótimos $x_1^*(\alpha_0), \dots, x_n^*(\alpha_0)$, para $\alpha = \alpha_0$, quando restrições adicionais são ajustadas ao problema sem alterar os valores $x_1^*(\alpha_0), \dots, x_n^*(\alpha_0)$. Isso significa perguntar-se sobre a relação entre $\frac{\partial^r x_i^*}{\partial \alpha_j}(\alpha_0)$ e $\frac{\partial^{r+1} x_i^*}{\partial \alpha_j}(\alpha_0)$, em que $^r x_1^*(\alpha_0), \dots, ^r x_n^*(\alpha_0)$ é a solução do problema:

$$\begin{aligned} \max_{x_1, \dots, x_n} \quad & f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = \phi_r(\alpha_1, \dots, \alpha_m) \\ \text{sa} \quad & g_1(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0 \\ & \vdots \\ & g_r(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0, \end{aligned}$$

e $^{r+1} x_1^*(\alpha_0), \dots, ^{r+1} x_n^*(\alpha_0)$ é a solução do problema:

$$\begin{aligned} \max_{x_1, \dots, x_n} \quad & f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = \phi_{r+1}(\alpha_1, \dots, \alpha_m) \\ \text{sa} \quad & g_1(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0 \\ & \vdots \\ & g_r(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0 \\ & g_{r+1}(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0, \end{aligned}$$

sendo $(x_1^*(\alpha_0), \dots, x_n^*(\alpha_0)) = (x_1^{r+1}(\alpha_0), \dots, x_n^{r+1}(\alpha_0))$. Observe que o comentário acima sobre longo e curto prazo é uma instância particular do problema geral.

Samuelson não dá muita atenção aos teoremas de 1ª ordem; a sua preocupação é estabelecer os teoremas de 2ª ordem, dos quais na seção “Restrições Auxiliares e o Princípio de Le Chatelier Generalizado” no Capítulo III do livro “Análise”.

Atualmente, sabemos que não foi Samuelson o primeiro a empregar um dos resultados de 1ª ordem: “no final da década de 1880, Rudolf Auspitz (1837-1906) e Richard Lieben (1842-1919) usaram o que é hoje conhecido como o teorema de envelope, apesar de não chamá-lo por este nome ou por qualquer outro nome¹¹.

Depois de Samuelson, trabalhos bastante citados são os artigos “*The Le Chatelier Principle as a Corollary to a Generalized Envelope Theorem*”¹² e “*A Revision of Comparative Statics Methodology in Economics, or, How to Do Comparative Statics on the Back of an Envelope*”¹³ (1972), ambos de Eugene Silberberg (1940-), nos quais introduz o chamado “problema primal-dual”.

3.1 Teorema 2: Teorema do Envelope (caso sem restrições)

Consideremos o problema:

$$\min/\max_{x_1, \dots, x_n} f(x_1, \dots, x_n, y_1, \dots, y_m) = V(y_1, \dots, y_m)$$

Então,

$$\frac{\partial V}{\partial y_j}(y) = \frac{\partial f}{\partial y_j}(x^*(y), y).$$

Seja $x^*(y) = (x_1^*(y), \dots, x_n^*(y))$ um ponto de mínimo ou máximo local de $f(x, y)$. Então, $V(y) = f(x^*(y), y)$. Pela regra da cadeia,

$$\frac{\partial V}{\partial y_j}(y) = \frac{\partial f}{\partial x_1}(x^*(y), y) \cdot \frac{\partial x_1^*}{\partial y_j}(y) + \dots + \frac{\partial f}{\partial x_n}(x^*(y), y) \cdot \frac{\partial x_n^*}{\partial y_j}(y) + \frac{\partial f}{\partial y_j}(x^*(y), y),$$

como $\frac{\partial f}{\partial x_1}(x^*(y), y) = \dots = \frac{\partial f}{\partial x_n}(x^*(y), y) = 0$, pois x^* é ótimo, então $\frac{\partial V}{\partial y_j}(y) = \frac{\partial f}{\partial y_j}(x^*(y), y)$.

Exemplo 5: Consideremos a função $f(x, y) = \frac{1}{(1 + x^2 - y^2 \cdot x)}$, em que y é um parâmetro exógeno e a maximização é feita sobre a variável x . Os máximos de $f(x, y)$ estão em $x = \frac{y^2}{2}$, fixado y . Portanto, $V(y) = \frac{4}{4 - y^4}$. Observe que $V'(y) = \frac{16 \cdot y^3}{(4 - y^4)^2} = \frac{\partial f}{\partial y} \left(\frac{y^2}{2}, y \right)$.

3.1.1 Lema de Hotelling

Dada a função de lucro para uma firma num mercado de livre competição:

$$\pi(p, w) = \max_{x_1, \dots, x_n} [p \cdot f(x_1, \dots, x_n) - w_1 \cdot x_1 - \dots - w_n \cdot x_n],$$

pelo Teorema do Envelope, $\frac{\partial \pi}{\partial p}(p, w) = f(x_1^*(p, w), \dots, x_n^*(p, w)) = y^*(p, w)$ e $\frac{\partial \pi}{\partial w_i}(p, w) = -x_i^*(p, w)$, que são resultados óbvios: se o preço aumentar R\$ 1 e $y^*(p, w)$ é o número de unidades produzidas, então o lucro aumenta $y^*(p, w)$; se o preço w_i do insumo i aumentar R\$ 1, o lucro $y^*(p, w)$ diminui $x_i^*(p, w)$. Esses resultados são conhecidos como “**Lemas de Hotelling**”¹⁴.

Exercício 1

Hotelling define a função de lucro da firma como $\pi(q_1, \dots, q_n) = \max_{p_1, \dots, p_n} [p_1 \cdot q_1 + \dots + p_n \cdot q_n - v(q_1, \dots, q_n)]$, em que $q_1, \dots, q_n$ são quantidades produzidas de diferentes produtos, $p_1, \dots, p_n$ os preços unitários dos produtos $i = 1, \dots, n$, respectivamente, e $v(q_1, \dots, q_n)$ é o custo de produção. Hotelling expressa a função lucro como função das quantidades (exógenas) e não dos preços (endógenos). Verifique que $\frac{\partial \pi}{\partial q_i}(q_1, \dots, q_n) = p_i^*(q_1, \dots, q_n) - \frac{\partial v}{\partial q_i}(q_1, \dots, q_n)$. Supondo livre competição, conclui que $\frac{\partial v}{\partial q_i}(q_1, \dots, q_n) = p_i^*(q_1, \dots, q_n)$.

Dados os preços $\bar{p}, \bar{w}$ temos uma demanda ótima $x_i^*(\bar{p}, \bar{w})$ para o insumo i . Fixada essa demanda e fixados os preços dos insumos em $\bar{w}$, temos duas funções de lucro. A função anterior:

$$\pi_L(p) = \max_{x_1, \dots, x_n} [p \cdot f(x_1, \dots, x_n) - \bar{w}_1 \cdot x_1 - \dots - \bar{w}_n \cdot x_n],$$

a que chamamos de lucro de longo prazo, na qual todas as variáveis $x_1, \dots, x_n$ são variáveis de escolha, e a função abaixo:

$$\pi_S(p) = \max_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} [p \cdot f(x_1, \dots, x_i^*(\bar{p}, \bar{w}), \dots, x_n) - \bar{w}_1 \cdot x_1 - \dots - \bar{w}_i \cdot x_i^*(\bar{p}, \bar{w}) - \dots - \bar{w}_n \cdot x_n],$$

na qual as variáveis de escolha são $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, ou seja, estamos excetuando x_i que está fixada em $x_i^*(\bar{p}, \bar{w})$. A essa última função $\pi_S(p)$ chamamos de lucro de curto prazo. Então temos as relações:

$$[\pi_L(p) - \pi_S(p)] \geq 0 \text{ e } [\pi_L(\bar{p}) - \pi_S(\bar{p})] = 0.$$

Portanto, $\bar{p}$ é um ponto de mínimo global para a função $[\pi_L(p) - \pi_S(p)]$. Logo, $\frac{d}{dp}[\pi_L(p) - \pi_S(p)]_{p=\bar{p}} = 0$ e $\frac{d^2}{dp^2}[\pi_L(p) - \pi_S(p)]_{p=\bar{p}} \geq 0$. Pelo Lema de Hotelling,

$$\frac{d}{dp}[\pi_L(p) - \pi_S(p)]_{p=\bar{p}} = y_L^*(p) - y_S^*(p), \quad \text{e,} \quad \text{portanto,} \quad y_L^*(\bar{p}) - y_S^*(\bar{p}) = 0. \quad \text{Ainda,}$$

$$\frac{d^2}{dp^2}[\pi_L(p) - \pi_S(p)]_{p=\bar{p}} = \frac{d}{dp}[y_L^*(p) - y_S^*(p)]_{p=\bar{p}} = \frac{d}{dp}[y_L^*(p)]_{p=\bar{p}} - \frac{d}{dp}[y_S^*(p)]_{p=\bar{p}} \geq 0.$$

Em outras palavras, $\frac{d}{dp}y_L^*(\bar{p}) \geq \frac{d}{dp}y_S^*(\bar{p})$, ou seja, o aumento da oferta de longo prazo é maior que o aumento da oferta de curto prazo em relação ao preço. Esse resultado ilustra o que chamamos “**Princípio de Le Chatelier-Samuelson**”.

3.2 Teorema 3: Teorema do Envelope (caso com restrições)

Consideremos o problema:

$$\begin{aligned} \min/\max_{x_1, \dots, x_n} \quad & f(x_1, \dots, x_n, y_1, \dots, y_m) = V(y_1, \dots, y_m) \\ \text{sa} \quad & g_1(x_1, \dots, x_n, y_1, \dots, y_m) \leq 0; \\ & \vdots \\ & g_k(x_1, \dots, x_n, y_1, \dots, y_m) \leq 0; \\ & h_1(x_1, \dots, x_n, y_1, \dots, y_m) \geq 0; \\ & \vdots \\ & h_l(x_1, \dots, x_n, y_1, \dots, y_m) \geq 0. \end{aligned}$$

A função $V(y_1, \dots, y_m) = V(y)$ é dita a função-valor e depende dos parâmetros exógenos $y = (y_1, \dots, y_m)$. Seja $x^*(y) = (x_1^*(y), \dots, x_n^*(y))$ um ponto de ótimo local do problema

acima e sejam $\lambda_1(y), \dots, \lambda_k(y), \mu_1(y), \dots, \mu_l(y)$ os respectivos multiplicadores de Lagrange que escrevemos como $\lambda_1^*, \dots, \lambda_k^*, \mu_1^*, \dots, \mu_l^*$. Então,

$$\begin{aligned} \frac{\partial V}{\partial y_j}(y) = & \frac{\partial f}{\partial y_j}(x^*(y), y) - \lambda_1^*(y) \cdot \frac{\partial g_1}{\partial y_j}(x^*(y), y) - \dots - \lambda_k^*(y) \cdot \frac{\partial g_k}{\partial y_j}(x^*(y), y) \\ & - \mu_1^*(y) \cdot \frac{\partial h_1}{\partial y_j}(x^*(y), y) - \dots - \mu_l^*(y) \cdot \frac{\partial h_l}{\partial y_j}(x^*(y), y). \end{aligned}$$

À primeira vista, a expressão acima não parece correta, pois os parâmetros $y_1, \dots, y_m$ não influenciam apenas a função objetivo $f(x_1, \dots, x_n, y_1, \dots, y_m)$, as restrições $g_i(x_1, \dots, x_n, y_1, \dots, y_m) \leq 0$, $h_j(x_1, \dots, x_n, y_1, \dots, y_m) \geq 0$ e os multiplicadores $\lambda_i^*(y), \mu_j^*(y)$, mas também as soluções ótimas $x_1^*(y), \dots, x_n^*(y)$ das variáveis de escolha. Presumivelmente, conjecturaríamos as derivadas $\frac{\partial x_i^*}{\partial y_j}$ nas expressões das derivadas de $V(y_1, \dots, y_m)$.

Seja $x^*(y) = (x_1^*(y), \dots, x_n^*(y))$ um ponto de mínimo ou de máximo local de $f(x, y)$ sujeita às restrições acima. Então, $V(y) = f(x^*(y), y)$. Pela regra da cadeia,

$$\begin{aligned} \frac{\partial V}{\partial y_j}(y) = & \frac{\partial f}{\partial x_1}(x^*(y), y) \cdot \frac{\partial x_1^*}{\partial y_j}(y) + \dots + \frac{\partial f}{\partial x_n}(x^*(y), y) \cdot \frac{\partial x_n^*}{\partial y_j}(y) + \frac{\partial f}{\partial y_j}(x^*(y), y) \\ \frac{\partial V}{\partial y_j}(y) = & \nabla_x f(x^*(y), y) \cdot \frac{\partial}{\partial y_j} x^*(y) + \frac{\partial f}{\partial y_j}(x^*(y), y). \end{aligned}$$

Observe que $\nabla_x f(x^*(y), y) = \left(\frac{\partial f}{\partial x_1}(x^*(y), y), \frac{\partial f}{\partial x_2}(x^*(y), y), \dots, \frac{\partial f}{\partial x_n}(x^*(y), y) \right)$ e $\frac{\partial}{\partial y_j} x^*(y) = \left(\frac{\partial x_1^*}{\partial y_j}(y), \frac{\partial x_2^*}{\partial y_j}(y), \dots, \frac{\partial x_n^*}{\partial y_j}(y) \right)$.

Mas, pela primeira condição de K-T,

$$\nabla_x f(x^*(y), y) = \sum \lambda_i^* \cdot \nabla_x g_i(x^*(y), y) + \sum \mu_j^* \cdot \nabla_x h_j(x^*(y), y),$$

logo, $\frac{\partial V}{\partial y_j}(y) =$

$$\begin{aligned} & \left[\sum \lambda_i^* \cdot \nabla_x g_i(x^*(y), y) + \sum \mu_j^* \cdot \nabla_x h_j(x^*(y), y) \right] \cdot \frac{\partial}{\partial y_j} x^*(y) + \frac{\partial f}{\partial y_j}(x^*(y), y), & \frac{\partial V}{\partial y_j}(y) = \\ & \sum \lambda_i^* \cdot \nabla_x g_i(x^*(y), y) \cdot \frac{\partial}{\partial y_j} x^*(y) + \sum \mu_j^* \cdot \nabla_x h_j(x^*(y), y) \cdot \frac{\partial}{\partial y_j} x^*(y) + \frac{\partial f}{\partial y_j}(x^*(y), y). \end{aligned}$$

Caso $\lambda_i^* \neq 0$ e $\mu_j^* \neq 0$, então $g_i(x^*(y), y) = 0$ e $h_j(x^*(y), y) = 0$, então

$$\nabla_x g_i(x^*(y), y) \cdot \frac{\partial}{\partial y_j} x^*(y) + \frac{\partial g_i}{\partial y_j}(x^*(y), y) = 0 \text{ e}$$

$$\nabla_x h_j(x^*(y), y) \cdot \frac{\partial}{\partial y_j} x^*(y) + \frac{\partial h_j}{\partial y_j}(x^*(y), y) = 0, \text{ pela regra da cadeia. Se } \lambda_i^* = 0 \text{ ou } \mu_j^* = 0 \text{ então}$$

$\lambda_i^* \cdot \nabla_x g_i(x^*(y), y)$ e $\mu_j^* \cdot \nabla_x h_j(x^*(y), y)$ não aparecem na expressão de $\frac{\partial V}{\partial y_j}(y)$.

Portanto, concluímos que $\frac{\partial V}{\partial y_j}(y) =$

$$\sum \lambda_i^*(y) \cdot \left(-\frac{\partial g_i}{\partial y_j}(x^*(y), y) \right) + \sum \mu_j^*(y) \cdot \left(-\frac{\partial h_j}{\partial y_j}(x^*(y), y) \right) + \frac{\partial f}{\partial y_j}(x^*(y), y).$$

Enfatizamos que os multiplicadores $\lambda_i^*(y)$ e $\mu_j^*(y)$ também dependem das variáveis exógenas $y = (y_1, \dots, y_m)$.

É comum apresentar as derivadas $\frac{\partial V}{\partial y_j}$ como derivadas do lagrangeano, mas é necessário certo cuidado. O lagrangeano do problema é

$$\mathcal{L}(x, y, \lambda, \mu) = f(x, y) - \sum \lambda_i(y) \cdot g_i(x, y) - \sum \mu_j(y) \cdot h_j(x, y).$$

Para apresentar $\frac{\partial V}{\partial y_j}(y)$ como $\frac{\partial \mathcal{L}}{\partial y_j}(x^*(y), y, \lambda^*, \mu^*) = \frac{\partial \mathcal{L}}{\partial y_j}(x^*, y, \lambda^*, \mu^*)$, devemos tomar os multiplicadores λ^* e μ^* como constantes no lagrangeano e não derivá-los em relação à y_j para obter a expressão correta:

$$\frac{\partial V}{\partial y_j}(y) = \sum \lambda_i^*(y) \cdot \left(-\frac{\partial g_i}{\partial y_j}(x^*(y), y) \right) + \sum \mu_j^*(y) \cdot \left(-\frac{\partial h_j}{\partial y_j}(x^*(y), y) \right) + \frac{\partial f}{\partial y_j}(x^*(y), y).$$

Exemplo 6: Consideremos o problema:

$$\begin{aligned} \min \quad & w_1 \cdot x_1 + w_2 \cdot x_2 = C(w_1, w_2, y) \\ \text{s.a.} \quad & f(x_1, x_2) \geq y \\ & x_1 > 0, x_2 > 0, \end{aligned}$$

tal que $w_1 > 0, w_2 > 0, \frac{\partial f}{\partial x_1} > 0$ e $\frac{\partial f}{\partial x_2} > 0$. Utilizando $\mathcal{L} = w_1 \cdot x_1 + w_2 \cdot x_2 - \mu \cdot (f(x_1, x_2) - y)$, concluímos, pelo Teorema do Envelope, $\frac{\partial C}{\partial w_1}(w_1, w_2, y) = x_1^*(w_1, w_2, y)$ e $\frac{\partial C}{\partial w_2}(w_1, w_2, y) = x_2^*(w_1, w_2, y)$. Esses dois resultados estão no livro “*Theory of Cost and Production Functions*”, de Ronald William Shephard (1912-1982), publicado em 1953¹⁵, e são chamados “Lemas de Shephard”. Mostre que $\frac{\partial C}{\partial y}(w_1, w_2, y) = \mu^*(w_1, w_2, y)$.

Exemplo 7: Consideremos o problema:

$$\begin{aligned} \max \quad & u(x_1, x_2) = u^*(p_1, p_2, m) \\ \text{s.a.} \quad & p_1 \cdot x_1 + p_2 \cdot x_2 = m \\ & x_1 > 0, x_2 > 0, \end{aligned}$$

tal que $p_1 > 0, p_2 > 0, \frac{\partial u}{\partial x_1} > 0$ e $\frac{\partial u}{\partial x_2} > 0$. Utilizando $\mathcal{L} = u(x_1, x_2) - \lambda \cdot (p_1 \cdot x_1 + p_2 \cdot x_2 - m)$, concluímos, pelo Teorema do Envelope, $\frac{\partial u^*}{\partial p_1}(p_1, p_2, m) = -\lambda^*(p_1, p_2, m) \cdot x_1^*(p_1, p_2, m)$, $\frac{\partial u^*}{\partial p_2}(p_1, p_2, m) = -\lambda^*(p_1, p_2, m) \cdot x_2^*(p_1, p_2, m)$ e $\frac{\partial u^*}{\partial m}(p_1, p_2, m) = \lambda^*(p_1, p_2, m)$. As chamadas “Identidades de Roy” são as expressões $x_1^*(p_1, p_2, m) = -\frac{\frac{\partial u^*}{\partial p_1}(p_1, p_2, m)}{\frac{\partial u^*}{\partial m}(p_1, p_2, m)}$ e $x_2^*(p_1, p_2, m) = -\frac{\frac{\partial u^*}{\partial p_2}(p_1, p_2, m)}{\frac{\partial u^*}{\partial m}(p_1, p_2, m)}$. Essas expressões receberam esse nome por aparecerem num artigo publicado, em 1947, pelo economista francês René François Joseph Roy (1894–1977).¹⁶

3.3 Envelopes, o teorema de Kuhn-Tucker e programação linear

Voltemos aos problemas lineares abaixo:

$$\begin{cases} \max & b' \cdot x \\ \text{sa} & A \cdot x \leq c \\ & x \geq 0 \end{cases} \text{ e } \begin{cases} \min & c' \cdot y \\ \text{sa} & A' \cdot y \geq b \\ & y \geq 0. \end{cases}$$

Na seção 7.2 da Lição Cinco, observamos que dois problemas têm o mesmo lagrangeano $\mathcal{L} = b' \cdot x - y' \cdot (A \cdot x - c) = c' \cdot y - x' \cdot (A' \cdot y - b)$. Também observamos que o multiplicador de Lagrange y^* associado à solução x^* satisfaz as restrições do problema dual e, igualmente, se y^* é solução do problema dual, então o multiplicador de Lagrange x^* associado à solução y^* satisfaz as restrições do problema primal. Além disso, verificamos que $b' \cdot x^* = c' \cdot y^*$.

Podemos combinar o Teorema do Envelope com a condição de Euler para mostrar que $b' \cdot x^* = c' \cdot y^*$.

Exercício 2

Uma função $f(x_1, \dots, x_n)$ é dita **positivamente homogênea** (homogênea) de grau k se $f(t \cdot x_1, \dots, t \cdot x_n) = t^k \cdot f(x_1, \dots, x_n)$, para $t > 0$. Não é difícil constatar que se f é homogênea de grau k , então satisfaz a **condição de Euler**

$$x_1 \cdot \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) + \dots + x_n \cdot \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) = k \cdot f(x_1, \dots, x_n).$$

i) Usando a regra da cadeia, obtenha $\frac{\partial}{\partial t} f(t \cdot x_1, \dots, t \cdot x_n) = x_1 \cdot \frac{\partial f}{\partial x_1}(t \cdot x_1, \dots, t \cdot x_n) + \dots + x_n \cdot \frac{\partial f}{\partial x_n}(t \cdot x_1, \dots, t \cdot x_n)$;

ii) Derive $\frac{\partial}{\partial t}(t^k \cdot f(x_1, \dots, x_n))$ para obter $k \cdot t^{k-1} \cdot f(x_1, \dots, x_n)$;

iii) Usando $f(t \cdot x_1, \dots, t \cdot x_n) = t^k \cdot f(x_1, \dots, x_n)$ e fazendo $t=1$, obtenha a condição de Euler.

O valor máximo no problema primal é:

$$b' \cdot x^* = b_1 \cdot x_1^* + b_2 \cdot x_2^* + \dots + b_n \cdot x_n^* = V(b_1, \dots, b_n, c_1, \dots, c_m).$$

Se fixarmos $b_1, \dots, b_n$, $V(b_1, \dots, b_n, c_1, \dots, c_m)$ é uma função homogênea de grau um em $c_1, \dots, c_m$, isto é, $V(b_1, \dots, b_n, t \cdot c_1, \dots, t \cdot c_m) = t \cdot V(b_1, \dots, b_n, c_1, \dots, c_m)$, $t > 0$ então, pela condição de Euler,

$$c_1 \cdot \frac{\partial V}{\partial c_1}(b_1, \dots, b_n, c_1, \dots, c_m) + \dots + c_m \cdot \frac{\partial V}{\partial c_m}(b_1, \dots, b_n, c_1, \dots, c_m) = V(b_1, \dots, b_n, c_1, \dots, c_m).$$

Mas, $\frac{\partial V}{\partial c_j}(b_1, \dots, b_n, c_1, \dots, c_m) = y_j^*$, pelo Teorema do Envelope, pois $\frac{\partial \mathcal{L}}{\partial c_j} = b' \cdot x - y' \cdot (A \cdot x - c) = y_j$.

Portanto,

$$c_1 \cdot y_1^* + \dots + c_m \cdot y_m^* = V(b_1, \dots, b_n, c_1, \dots, c_m) = b_1 \cdot x_1^* + b_2 \cdot x_2^* + \dots + b_n \cdot x_n^*.$$

Exercício 3

Observe que $V(b_1, \dots, b_n, c_1, \dots, c_m)$ é uma função homogênea de grau um em $c_1, \dots, c_m$, pois x^* é a solução do problema $\begin{cases} \max & b' \cdot x \\ \text{sa} & A \cdot x \leq c \\ & x \geq 0 \end{cases}$ se e só se $t \cdot x^*$ é solução do problema $\begin{cases} \max & b' \cdot x \\ \text{sa} & A \cdot x \leq t \cdot c \\ & x \geq 0 \end{cases}$ para $t > 0$.

A seguir, responderemos uma pergunta feita na Lição Três em relação ao Método Simplex: por que os coeficientes das variáveis de folga na equação $F = \dots - y_k^* \cdot x_{n+k} - \dots$ são os valores ótimos do problema dual?

Consideremos o problema:

$$\begin{aligned} \max_{x_1, \dots, x_n} \quad & b' \cdot x & = F(b, c, x_{n+1}, \dots, x_{n+m}) \\ \text{sa} \quad & A \cdot x + z = c \\ & x \geq 0, z \geq 0, \end{aligned}$$

em que $z = \begin{pmatrix} x_{n+1} \\ \vdots \\ x_{n+m} \end{pmatrix} \geq 0$ é o vetor das variáveis de folga. O lagrangeano do problema acima é $\mathcal{L} = b_1 \cdot x_1 + \dots + b_n \cdot x_n - \lambda_{n+1} \cdot (a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n + x_{n+1} - c_1) - \dots$

$$-\lambda_{n+m} \cdot (a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n + x_{n+m} - c_1).$$

Pelo Teorema do Envelope, $\frac{\partial F}{\partial x_{n+k}}(b, c, x_{n+1}, \dots, x_{n+m}) = -\lambda_{n+k}^*$. Pela dualidade, $\lambda_{n+k}^* = y_k^*$ em que $y_1^*, \dots, y_m^*$ são as soluções do problema dual $\max_{y_1, \dots, y_m} c' \cdot y$
 $s.a. \quad A' \cdot y \geq a.$
 $y \geq 0,$

Pelo *tableau* simplex, se $F = \dots - \theta_k^* \cdot x_{n+k} - \dots$ e, portanto, $\frac{\partial F}{\partial x_{n+k}} = \frac{\partial V}{\partial x_{n+k}} = -\theta_k^* = -\lambda_k^* = -y_k^*$. Observe no *tableau* simplex:

	...	x_{n+k}	...
$\vdots$			
x_j			x_j^*
$\vdots$			
F	...	θ_k^*	...

Logo, os valores ótimos no problema dual são os coeficientes das variáveis de folga na última linha do *Tableau* Simplex.

Exercício 4

Considere o problema:

$$\begin{aligned} \min \quad & w_1 \cdot x_1 + \dots + w_n \cdot x_n = C(w_1, \dots, w_n, y) \\ \text{s.a.} \quad & f(x_1, \dots, x_n) \geq y \\ & x_1 > 0, \dots, x_n > 0. \end{aligned}$$

Mostre que $\frac{\partial C}{\partial w_j}(w_1, \dots, w_n, y) = x_j^*(w_1, \dots, w_n, y)$ e que $\frac{\partial C}{\partial y}(w_1, w_2, y) = \mu^*(w_1, w_2, y)$. Derivando $C(w_1, \dots, w_n, y) = w_1 \cdot x_1^*(w_1, \dots, w_n, y) + \dots + w_n \cdot x_n^*(w_1, \dots, w_n, y)$ em relação a w_i e usando os lemas de Shephard, conclua que $w_1 \cdot \frac{\partial x_1^*}{\partial w_j} + \dots + w_n \cdot \frac{\partial x_n^*}{\partial w_j} = 0, j = 1, \dots, n$. Esse resultado também pode ser obtido observando que $C(t \cdot w_1, \dots, t \cdot w_n, y) = t \cdot C(w_1, \dots, w_n, y), t > 0$. Por que?

Exercício 5

Considere o problema $\max f(x_1, x_2) + a \cdot x_1 = V(a, b)$ s.a. $g(x_1, x_2) + b \cdot x_2 = 0$. Usando o Teorema do Envelope, obtenha $\frac{\partial V}{\partial a}(a, b)$ e $\frac{\partial V}{\partial b}(a, b)$.

Exercício 6

Considere o problema $\max_{x_1, \dots, x_n} f(x_1, x_2) + h(x_1, a) = V(a, b)$ s.a. $g(x_1, x_2, b) = 0$. Usando o Teorema do Envelope, obtenha $\frac{\partial V}{\partial a}(a, b)$ e $\frac{\partial V}{\partial b}(a, b)$.

4. Estática comparativa

Seja o problema $\pi(p, w) = \max_{x_1, \dots, x_n} [p \cdot f(x_1, \dots, x_n) - w_1 \cdot x_1 - \dots - w_n \cdot x_n]$. As condições de primeira ordem são $\frac{\partial f}{\partial x_1}(x_1^*(p, w), \dots, x_n^*(p, w)) - \frac{w_1}{p} = 0$, $\frac{\partial f}{\partial x_2}(x_1^*(p, w), \dots, x_n^*(p, w)) - \frac{w_2}{p} = 0, \dots, \frac{\partial f}{\partial x_n}(x_1^*(p, w), \dots, x_n^*(p, w)) - \frac{w_n}{p} = 0$.

Devemos nos assegurar que essas equações tenham solução. Pelo Teorema

da Função Implícita, se
$$\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2}(x, p, w) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x, p, w) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x, p, w) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x, p, w) \end{vmatrix} \neq 0$$
, podemos expressar $x_1, x_2, \dots, x_n$

em função de (p, w) localmente, ou seja, existe uma vizinhança de (p, w) na qual as funções $x_1^*(p, w), \dots, x_n^*(p, w)$ estão definidas. A fórmula $\partial_x y(x) = -(\partial_y f(x, y(x)))^{-1} \cdot (\partial_x f(x, y(x)))$ traduz-se em $\partial_{p, w} x^*(p, w) = -(\partial_x F(x^*(p, w), p, w))^{-1} \cdot (\partial_{p, w} F(x^*(p, w), p, w))$, em que $F: \square^{n+(n+1)} \rightarrow \square^n$ é $F(x, p, w) = (F_1(x, p, w), \dots, F_n(x, p, w))$, $F_i(x, p, w) = \frac{\partial f}{\partial x_i}(x) - \frac{w_i}{p} = 0$.

Então,

$$\begin{pmatrix} \frac{\partial x_1^*}{\partial p} & \frac{\partial x_1^*}{\partial w_1} & \dots & \frac{\partial x_1^*}{\partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n^*}{\partial p} & \frac{\partial x_n^*}{\partial w_1} & \dots & \frac{\partial x_n^*}{\partial w_n} \end{pmatrix} = - \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x^*) \end{pmatrix}^{-1} \begin{pmatrix} \frac{w_1}{p^2} & -\frac{1}{p} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{p^2} & 0 & 0 & \dots & -\frac{1}{p} \end{pmatrix}$$

ou ainda,

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x^*) \end{pmatrix} \begin{pmatrix} \frac{\partial x_1^*}{\partial p} & \frac{\partial x_1^*}{\partial w_1} & \dots & \frac{\partial x_1^*}{\partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n^*}{\partial p} & \frac{\partial x_n^*}{\partial w_1} & \dots & \frac{\partial x_n^*}{\partial w_n} \end{pmatrix} = - \begin{pmatrix} \frac{w_1}{p^2} & -\frac{1}{p} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{p^2} & 0 & 0 & \dots & -\frac{1}{p} \end{pmatrix}.$$

Por exemplo,

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*(p, w)) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*(p, w)) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*(p, w)) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x^*(p, w)) \end{pmatrix} \begin{pmatrix} \frac{\partial x_1^*}{\partial w_i} \\ \vdots \\ \frac{\partial x_n^*}{\partial w_i} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \frac{1}{p} \\ \vdots \\ 0 \end{pmatrix} \leftarrow i,$$

que podemos resolver por Cramer:

$$\frac{\partial x_j^*}{\partial w_i} = \frac{\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*(p, w)) & \cdots & 0 & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*(p, w)) \\ \vdots & & \downarrow j & & \vdots \\ \vdots & & \frac{1}{p} & \leftarrow i & \vdots \\ \vdots & & \vdots & & \vdots \end{vmatrix}}{\begin{vmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*(p, w)) & \cdots & 0 & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x^*(p, w)) \\ \frac{\partial^2 f}{\partial x_1^2}(x^*(p, w)) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*(p, w)) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*(p, w)) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x^*(p, w)) \end{vmatrix}},$$

$\frac{\partial x_j^*}{\partial w_i} = \frac{(-1)^{i+j} \cdot \frac{1}{p} \cdot |f''(x^*(p, w))(i|j)|}{|f''(x^*(p, w))|}$, em que $|f''(x^*(p, w))(i|j)|$ resulta de eliminar a linha i e a coluna j na matriz $f''(x^*(p, w))$.

Se supusermos $f''(x^*(p, w))$ negativa-definida, então $(-1)^n \cdot |f''(x^*(p, w))| > 0$ e $(-1)^{n-1} \cdot |f''(x^*(p, w))(i|i)| > 0$, ou seja, $|f''(x^*(p, w))|$ e $|f''(x^*(p, w))(i|i)|$ têm sinais opostos, então $\frac{\partial x_j^*}{\partial w_i} = \frac{(-1)^{i+j} \cdot \frac{1}{p} \cdot |f''(x^*(p, w))(i|j)|}{|f''(x^*(p, w))|} = \frac{|f''(x^*(p, w))(i|j)|}{p \cdot |f''(x^*(p, w))|} < 0$. Como $|f''(x^*(p, w))(i|j)| = |f''(x^*(p, w))(j|i)|$, porque $f''(x^*(p, w))$ é simétrica, então $\frac{\partial x_j^*}{\partial w_i} = \frac{(-1)^{i+j} \cdot \frac{1}{p} \cdot |f''(x^*(p, w))(i|j)|}{|f''(x^*(p, w))|} = \frac{(-1)^{i+j} \cdot \frac{1}{p} \cdot |f''(x^*(p, w))(j|i)|}{|f''(x^*(p, w))|} = \frac{\partial x_i^*}{\partial w_j}$, que é chamada de Lei de Reciprocidade. Observe que postular $f''(x^*(p, w))$ negativa-definida implica que $x^*(p, w)$ é um mínimo global.

Exercício 7

No problema $\pi(p, w) = \max_{x_1, x_2} [p \cdot f(x_1, x_2) - w_1 \cdot x_1 - w_2 \cdot x_2]$, mostre, fazendo os cálculos para esse caso particular, que $\frac{\partial x_1^*}{\partial w_1} = \frac{f_{22}}{p \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})} < 0$, $\frac{\partial x_2^*}{\partial w_2} = \frac{f_{11}}{p \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})} < 0$, $\frac{\partial x_1^*}{\partial w_2} = \frac{-f_{12}}{p \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})}$ e $\frac{\partial x_2^*}{\partial w_1} = \frac{-f_{21}}{p \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})}$. Facilmente vemos que $\frac{\partial x_1^*}{\partial w_2} = \frac{\partial x_2^*}{\partial w_1}$.

Exercício 8

Suponhamos que fixamos $x_2 = x_2^*(p, w_1, w_2)$ e consideremos o problema $\pi(p, w) = \max_{x_1} [p \cdot f(x_1, x_2^*) - w_1 \cdot x_1 - w_2 \cdot x_2^*]$. A condição de primeira ordem é $\frac{\partial f}{\partial x_1}(x_1^*, x_2^*) = \frac{w_1}{p}$, em que $x_1^*(p, w_1, w_2)$ é dita a demanda de curto prazo pelo insumo um, pois a demanda pelo insumo dois está fixada em $x_2^*(p, w_1, w_2)$. As demandas $x_1^*(p, w_1, w_2)$ e $x_2^*(p, w_1, w_2)$ são ditas de longo prazo. Mostre que $\frac{\partial x_1^{s*}}{\partial w_1} = \frac{1}{p \cdot f_{11}} < 0$ e $\frac{\partial x_1^*}{\partial w_1} - \frac{\partial x_1^{s*}}{\partial w_1} = \frac{(f_{12})^2}{p \cdot f_{11} \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})} \leq 0$. Mostre que $\frac{\partial x_2^{s*}}{\partial w_2} = \frac{1}{p \cdot f_{22}} < 0$ e $\frac{\partial x_2^*}{\partial w_2} - \frac{\partial x_2^{s*}}{\partial w_2} = \frac{(f_{12})^2}{p \cdot f_{22} \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})} \leq 0$. Observe que $\frac{\partial x_1^*}{\partial w_1} \cdot \frac{w_1}{x_1^*} - \frac{\partial x_1^{s*}}{\partial w_1} \cdot \frac{w_1}{x_1^*} = -\varepsilon_{11} + \varepsilon_{11}^s \leq 0$, ou seja, a elasticidade da demanda de longo prazo pelo insumo um em relação ao

preço w_1 é naturalmente maior que a elasticidade dessa demanda no curto prazo.

Exercício 9

Observe que
$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x^*) \end{pmatrix} \begin{pmatrix} \frac{\partial x_1^*}{\partial p} \\ \vdots \\ \frac{\partial x_n^*}{\partial p} \end{pmatrix} = \begin{pmatrix} -\frac{w_1}{p^2} \\ \vdots \\ -\frac{w_n}{p^2} \end{pmatrix}.$$
 Mostre que no caso $n=2$,
$$\frac{\partial x_1^*}{\partial p} = -\frac{(w_1 \cdot f_{22} - w_2 \cdot f_{12})}{p^2 \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})} \text{ e } \frac{\partial x_2^*}{\partial p} = -\frac{(w_2 \cdot f_{11} - w_1 \cdot f_{21})}{p^2 \cdot (f_{11} \cdot f_{22} - f_{12} \cdot f_{21})}.$$

4.1 A Metodologia primal-dual de Eugene Silberberg¹⁷ (caso sem restrições)

Vamos usar o problema tratado acima para expor a metodologia de Silberberg para Estática Comparativa no caso não-restrito. Consideremos o problema de maximização do lucro:

$$\pi(w, p) = \max_{x_1, \dots, x_n} [p \cdot f(x_1, \dots, x_n) - w_1 \cdot x_1 - \dots - w_n \cdot x_n].$$

Sejam $x_1^*(w, p)$, $x_2^*(w, p)$, ..., $x_n^*(w, p)$ as demandas por insumos. Notemos $x = (x_1, \dots, x_n)$ e $w = (w_1, \dots, w_n)$. Observe que a função $F(x, w, p) = \pi(w, p) - p \cdot f(x_1, \dots, x_n) + w_1 \cdot x_1 + \dots + w_n \cdot x_n \geq 0$ tem um mínimo em $(x^*(w, p), w, p)$, pois $F(x^*(w, p), w, p) = 0$. Essa é a exigência que fazemos nesse problema. O problema não tem sentido se não impusermos a condição que os pontos críticos de $F(x, w, p)$ sejam mínimos locais. Fixado (w, p) , a exigência é que $x^*(w, p)$ seja um máximo global de $[p \cdot f(x_1, \dots, x_n) - w_1 \cdot x_1 - \dots - w_n \cdot x_n]$, mas como, em relação à função $F(x, w, p)$, cada $(x^*(w, p), w, p)$ é um mínimo, então existem infinitos mínimos para $F(x, w, p)$. Nesse caso, a exigência mínimo global para $F(x, w, p)$ tem um significado diferente daquele que queremos. Portanto, exijamos $F''(x^*, w, p)$ positiva-semidefinida. As condições de 1ª ordem são $\frac{\partial F}{\partial x_1} = -p \cdot \frac{\partial f}{\partial x_1} + w_1 = 0, \dots, \frac{\partial F}{\partial x_n} = -p \cdot \frac{\partial f}{\partial x_n} + w_n = 0$, $\frac{\partial F}{\partial w_1} = \frac{\partial \pi}{\partial w_1} + x_1 = 0, \dots, \frac{\partial F}{\partial w_n} = \frac{\partial \pi}{\partial w_n} + x_n = 0$ e $\frac{\partial F}{\partial p} = \frac{\partial \pi}{\partial p} - f = 0$ (essa última igualdade é o Teorema do Envelope); a condição de 2ª ordem implica que os menores principais da

$$F''(x^*, w, p) = \begin{pmatrix} -p \cdot f_{11} & \cdots & -p \cdot f_{1n} & 1 & \cdots & 0 & -f_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -p \cdot f_{n1} & \cdots & -p \cdot f_{nn} & 0 & \cdots & 1 & -f_n \\ 1 & \cdots & 0 & \pi_{w_1 w_1} & \cdots & \pi_{w_1 w_n} & \pi_{w_1 p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & \pi_{w_n w_1} & \cdots & \pi_{w_n w_n} & \pi_{w_n p} \\ -f_1 & \cdots & -f_n & \pi_{pw_1} & \cdots & \pi_{pw_n} & \pi_{pp} \end{pmatrix}$$

são maiores ou iguais a zero. Portanto, $\begin{pmatrix} \pi_{w_1 w_1} & \cdots & \pi_{w_1 w_n} & \pi_{w_1 p} \\ \vdots & \ddots & \vdots & \vdots \\ \pi_{w_n w_1} & \cdots & \pi_{w_n w_n} & \pi_{w_n p} \\ \pi_{pw_1} & \cdots & \pi_{pw_n} & \pi_{pp} \end{pmatrix}$ é positiva-semidefinida. Logo, concluímos que a função lucro $\pi(w, p) = p \cdot f(x_1^*(w, p), \dots, x_n^*(w, p)) - w_1 \cdot x_1^*(w, p) - \dots - w_n \cdot x_n^*(w, p)$ é uma função convexa.

Observe que $\begin{pmatrix} -p \cdot f_{11}(x^*(w, p)) & \cdots & -p \cdot f_{1n}(x^*(w, p)) \\ \vdots & \ddots & \vdots \\ -p \cdot f_{n1}(x^*(w, p)) & \cdots & -p \cdot f_{nn}(x^*(w, p)) \end{pmatrix}$ é positiva-semidefinida, em que $x^*(w, p) = (x_1^*(w, p), \dots, x_n^*(w, p))$, portanto, um menor principal de ordem k $\begin{vmatrix} -p \cdot f_{h,h}(x^*(w, p)) & \cdots & -p \cdot f_{h,k}(x^*(w, p)) \\ \vdots & \ddots & \vdots \\ -p \cdot f_{k,h}(x^*(w, p)) & \cdots & -p \cdot f_{k,k}(x^*(w, p)) \end{vmatrix} = (-1)^k \cdot p^k \cdot \begin{vmatrix} f_{h,h}(x^*(w, p)) & \cdots & f_{h,k}(x^*(w, p)) \\ \vdots & \ddots & \vdots \\ f_{k,h}(x^*(w, p)) & \cdots & f_{k,k}(x^*(w, p)) \end{vmatrix} \geq 0$, o que é compatível com $x^*(w, p)$ ser máximo local de f .

Pelo Teorema do Envelope, $\pi_{w_i} = -x_i^*(w, p)$, portanto, $\pi_{w_i w_j} = -\frac{\partial x_i^*}{\partial w_j}(w, p)$ e como $\pi_{w_i w_j} = \pi_{w_j w_i}$, então $\frac{\partial x_i^*}{\partial w_j}(w, p) = \frac{\partial x_j^*}{\partial w_i}(w, p)$. Ainda pelo Teorema do Envelope, $\pi_p = f(x_1^*(w, p), \dots, x_n^*(w, p)) = f^*$, que é a quantidade produzida. Substituindo no determinante acima, obtemos

$$\begin{pmatrix} -\frac{\partial x_1^*}{\partial w_1} & \cdots & -\frac{\partial x_1^*}{\partial w_n} & -\frac{\partial x_1^*}{\partial p} \\ \vdots & \ddots & \vdots & \vdots \\ -\frac{\partial x_n^*}{\partial w_1} & \cdots & -\frac{\partial x_n^*}{\partial w_n} & -\frac{\partial x_n^*}{\partial p} \\ -\frac{\partial x_1^*}{\partial p} & \cdots & -\frac{\partial x_n^*}{\partial p} & \frac{\partial f^*}{\partial p} \end{pmatrix}, \text{ que é positiva-semidefinida, logo } \frac{\partial x_1^*}{\partial w_1} \leq 0, \dots, \frac{\partial x_n^*}{\partial w_n} \leq 0 \text{ e } \frac{\partial f^*}{\partial p} \geq 0.$$

A metodologia de Silberberg consiste em tratar variáveis endógenas e exógenas num mesmo problema de otimização substituindo o problema inicial por um outro em que a função valor e a função objetivo sejam otimizadas conjuntamente. As variáveis de escolha, então, são as variáveis iniciais da função objetivo e os parâmetros exógenos. Se o problema inicial é $V(\alpha_1, \dots, \alpha_m) = \max_{x_1, \dots, x_n} f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m)$, então o problema “primal-dual” torna-se $\min_{x_1, \dots, x_n, \alpha_1, \dots, \alpha_m} [V(\alpha_1, \dots, \alpha_m) - f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m)]$. Observe que o Teorema do Envelope resulta das condições de 1ª ordem desse problema, pois

$$\frac{\partial}{\partial \alpha_i} [V(\alpha_1, \dots, \alpha_m) - f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m)] = \frac{\partial V}{\partial \alpha_i} - \frac{\partial}{\partial \alpha_i} f(x_1^*, \dots, x_n^*, \alpha_1, \dots, \alpha_m) = 0.$$

Considere o ponto ótimo $(x_1^*(\alpha), \dots, x_n^*(\alpha), \alpha)$ do problema $\min_{x_1, \dots, x_n, \alpha_1, \dots, \alpha_m} [V(\alpha_1, \dots, \alpha_m) - f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m)]$. As condições de 1ª ordem são $-f_1 = 0, \dots, -f_n = 0, V_{\alpha_1} - f_{\alpha_1} = 0, \dots, V_{\alpha_m} - f_{\alpha_m} = 0$. No ponto ótimo, a forma quadrática

$$(V - f)'' = \begin{pmatrix} -f_{11} & \cdots & -f_{1n} & -f_{1\alpha_1} & \cdots & -f_{1\alpha_m} \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ -f_{n1} & \cdots & -f_{nn} & -f_{n\alpha_1} & \cdots & -f_{n\alpha_m} \\ -f_{\alpha_1 1} & \cdots & -f_{\alpha_1 n} & (V_{\alpha_1 \alpha_1} - f_{\alpha_1 \alpha_1}) & \cdots & (V_{\alpha_1 \alpha_m} - f_{\alpha_1 \alpha_m}) \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ -f_{\alpha_m 1} & \cdots & -f_{\alpha_m n} & (V_{\alpha_m \alpha_1} - f_{\alpha_m \alpha_1}) & \cdots & (V_{\alpha_m \alpha_m} - f_{\alpha_m \alpha_m}) \end{pmatrix}$$

é positiva-semidefinida, portanto, os

elementos da diagonal principal da matriz acima são tais que $(V_{\alpha_i \alpha_i}(\alpha) - f_{\alpha_i \alpha_i}(x_1^*(\alpha), \dots, x_n^*(\alpha), \alpha)) \geq 0$.

A partir das condições de 1ª ordem $V_{\alpha_i}(\alpha) - f_{\alpha_i}(x_1^*(\alpha), \dots, x_n^*(\alpha), \alpha) = 0$, concluímos que $V_{\alpha_i \alpha_i} - \left(f_{\alpha_i x_1} \cdot \frac{\partial x_1^*}{\partial \alpha_i} + \dots + f_{\alpha_i x_n} \cdot \frac{\partial x_n^*}{\partial \alpha_i} + f_{\alpha_i \alpha_i} \right) = 0$, ou seja,

$$V_{\alpha_i \alpha_i} - f_{\alpha_i \alpha_i} = f_{\alpha_i x_1} \cdot \frac{\partial x_1^*}{\partial \alpha_i} + \dots + f_{\alpha_i x_n} \cdot \frac{\partial x_n^*}{\partial \alpha_i} \geq 0, i = 1, \dots, m.$$

Silberberg observa que a relação acima foi obtida por Samuelson no livro “Fundamentos da Análise Econômica”, mas assinala que a matriz $(V - f)''$ oferece um conjunto mais amplo de relações para serem exploradas, isto é, todos os menores principais são não-negativos.

As “relações recíprocas” $f_{\alpha_i x_1} \cdot \frac{\partial x_1^*}{\partial \alpha_j} + \dots + f_{\alpha_i x_n} \cdot \frac{\partial x_n^*}{\partial \alpha_j} = f_{\alpha_j x_1} \cdot \frac{\partial x_1^*}{\partial \alpha_i} + \dots + f_{\alpha_j x_n} \cdot \frac{\partial x_n^*}{\partial \alpha_i}$ são obtidas através de $V_{\alpha_i \alpha_j} = V_{\alpha_j \alpha_i}$, em que $V_{\alpha_i \alpha_j} = f_{\alpha_i x_1} \cdot \frac{\partial x_1^*}{\partial \alpha_j} + \dots + f_{\alpha_i x_n} \cdot \frac{\partial x_n^*}{\partial \alpha_j} + f_{\alpha_i \alpha_j}$ e $V_{\alpha_j \alpha_i} = f_{\alpha_j x_1} \cdot \frac{\partial x_1^*}{\partial \alpha_i} + \dots + f_{\alpha_j x_n} \cdot \frac{\partial x_n^*}{\partial \alpha_i} + f_{\alpha_j \alpha_i}$.

4.3 Minimização de custos

Consideremos o problema:

$$\begin{aligned} \min \quad & w_1 \cdot x_1 + w_2 \cdot x_2 = C(w_1, w_2, y) \\ \text{s.a.} \quad & f(x_1, x_2) \geq y \\ & x_1 > 0, x_2 > 0, \end{aligned}$$

tal que $w_1 > 0, w_2 > 0, \frac{\partial f}{\partial x_1} > 0$ e $\frac{\partial f}{\partial x_2} > 0$. Se $\mathcal{L} = w_1 \cdot x_1 + w_2 \cdot x_2 - \mu \cdot (f(x_1, x_2) - y)$, as condições de 1ª ordem são: $\frac{\partial \mathcal{L}}{\partial x_1} = w_1 - \mu \cdot \frac{\partial f}{\partial x_1} = 0, \frac{\partial \mathcal{L}}{\partial x_2} = w_2 - \mu \cdot \frac{\partial f}{\partial x_2} = 0$ e $\frac{\partial \mathcal{L}}{\partial \mu} = -f + y = 0$. Observe que nas três equações temos seis variáveis: w_1, w_2, y , que são exógenas, e x_1, x_2, μ , que são endógenas.

Lembre-se que $\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & 0 \end{vmatrix} = D_2 > 0$ é suficiente para assegurar que f é quase-

côncava. Observe que

$$\begin{vmatrix} \frac{\partial^2 \mathcal{L}}{\partial x_1^2} & \frac{\partial^2 \mathcal{L}}{\partial x_2 \partial x_1} & \frac{\partial^2 \mathcal{L}}{\partial \mu \partial x_1} \\ \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_2} & \frac{\partial^2 \mathcal{L}}{\partial x_2^2} & \frac{\partial^2 \mathcal{L}}{\partial \mu \partial x_2} \\ \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial \mu} & \frac{\partial^2 \mathcal{L}}{\partial x_2 \partial \mu} & \frac{\partial^2 \mathcal{L}}{\partial \mu^2} \end{vmatrix} = -\mu \cdot \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & 0 \end{vmatrix} = -\mu \cdot D_2 < 0,$$

então, pelo teorema da função implícita, é possível expressar x_1, x_2, μ em função de w_1, w_2, y . Além disso, os pontos que satisfizerem as condições de Kuhn-Tucker para mínimos são mínimos globais.

Derivando as condições de 1ª ordem em relação a w_1 ,

$$\begin{pmatrix} -\mu \cdot \frac{\partial^2 f}{\partial x_1^2} & -\mu \cdot \frac{\partial^2 f}{\partial x_2 \partial x_1} & -\frac{\partial f}{\partial x_1} \\ -\mu \cdot \frac{\partial^2 f}{\partial x_1 \partial x_2} & -\mu \cdot \frac{\partial^2 f}{\partial x_2^2} & -\frac{\partial f}{\partial x_2} \\ -\frac{\partial f}{\partial x_1} & -\frac{\partial f}{\partial x_2} & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial x_1^*}{\partial w_1} \\ \frac{\partial x_2^*}{\partial w_1} \\ \frac{\partial \mu^*}{\partial w_1} \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}.$$

Usando Cramer,

$$\frac{\partial x_1^*}{\partial w_1} = \frac{\begin{vmatrix} -1 & -\mu \cdot f_{12} & -f_1 \\ 0 & -\mu \cdot f_{22} & -f_2 \\ 0 & -f_2 & 0 \end{vmatrix}}{\begin{vmatrix} -\mu^* \cdot f_{11}^* & -\mu^* \cdot f_{12}^* & -f_1^* \\ -\mu \cdot f_{21} & -\mu \cdot f_{22} & -f_2^* \\ -f_1^* & -f_2^* & 0 \end{vmatrix}} = \frac{f_2^2}{-\mu \cdot D_2} < 0,$$

$$\frac{\partial x_2^*}{\partial w_1} = \frac{\begin{vmatrix} -\mu \cdot f_{11} & -1 & -f_1 \\ -\mu \cdot f_{21} & 0 & -f_2 \\ -f_1 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} -\mu^* \cdot f_{11}^* & -\mu^* \cdot f_{12}^* & -f_1^* \\ -\mu \cdot f_{21} & -\mu \cdot f_{22} & -f_2^* \\ -f_1^* & -f_2^* & 0 \end{vmatrix}} = \frac{f_1 \cdot f_2}{\mu \cdot D_2} > 0,$$

$$\frac{\partial \mu^*}{\partial w_1} = \frac{\begin{vmatrix} -\mu^* \cdot f_{11}^* & -\mu^* \cdot f_{12}^* & -1 \\ -\mu^* \cdot f_{21}^* & -\mu^* \cdot f_{22}^* & 0 \\ -f_1^* & -f_2^* & 0 \end{vmatrix}}{\begin{vmatrix} -\mu^* \cdot f_{11}^* & -\mu^* \cdot f_{12}^* & -f_1^* \\ -\mu \cdot f_{21} & -\mu \cdot f_{22} & -f_2^* \\ -f_1^* & -f_2^* & 0 \end{vmatrix}} = -\frac{\begin{vmatrix} f_{21} & f_{22} \\ f_1 & f_2 \end{vmatrix}}{D_2}.$$

As derivadas $\frac{\partial x_1^*}{\partial w_2}$, $\frac{\partial x_2^*}{\partial w_2}$ e $\frac{\partial \mu^*}{\partial w_2}$ são obtidas de maneira similar.

Exemplo 8: Considere as derivadas das condições de 1ª ordem em relação a y :

$$\begin{pmatrix} -\mu^* \cdot \frac{\partial^2 f}{\partial x_1^2} & -\mu^* \cdot \frac{\partial^2 f}{\partial x_2 \partial x_1} & -\frac{\partial f}{\partial x_1} \\ -\mu^* \cdot \frac{\partial^2 f}{\partial x_1 \partial x_2} & -\mu^* \cdot \frac{\partial^2 f}{\partial x_2^2} & -\frac{\partial f}{\partial x_2} \\ -\frac{\partial f}{\partial x_1} & -\frac{\partial f}{\partial x_2} & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial x_1^*}{\partial y} \\ \frac{\partial x_2^*}{\partial y} \\ \frac{\partial \mu^*}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix},$$

então,

$$\frac{\partial \mu^*}{\partial y} = \frac{\begin{vmatrix} -\mu^* \cdot \frac{\partial^2 f}{\partial x_1^2} & -\mu^* \cdot \frac{\partial^2 f}{\partial x_2 \partial x_1} & 0 \\ -\mu^* \cdot \frac{\partial^2 f}{\partial x_1 \partial x_2} & -\mu^* \cdot \frac{\partial^2 f}{\partial x_2^2} & 0 \\ -\frac{\partial f}{\partial x_1} & -\frac{\partial f}{\partial x_2} & -1 \end{vmatrix}}{-\mu^* \cdot D_2} = \frac{\mu^* \cdot \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix}}{D_2} = \frac{\mu^* \cdot |f''|}{D_2}.$$

O sinal de $\frac{\partial \mu^*}{\partial y} = \frac{\mu^* \cdot |f''|}{D_2}$ não está definido, pois podemos ter $\frac{|f''|}{D_2} < 0$, $\frac{|f''|}{D_2} = 0$ ou $\frac{|f''|}{D_2} > 0$.

Exercício 10

Mostre que se $f(x, y) = x \cdot y$, $f(x, y) = x^{1/2} \cdot y^{1/2}$ e $f(x, y) = \ln(x+1) + \ln(y+1)$, então $\frac{|f''|}{D_2} < 0$, $\frac{|f''|}{D_2} = 0$ e $\frac{|f''|}{D_2} > 0$, respectivamente. Observe que as três funções são quase-côncavas, pois $D_2 > 0$, mas, a primeira função é quase-côncava sem ser côncava, a segunda é côncava não-estrita e a terceira é estritamente côncava.

Exercício 11

Obtenha as derivadas $\frac{\partial x_1^*}{\partial w_2}$, $\frac{\partial x_2^*}{\partial w_2}$ e $\frac{\partial \mu^*}{\partial w_2}$.

Exercício 12

Obtenha as derivadas $\frac{\partial x_1^*}{\partial y}$, $\frac{\partial x_2^*}{\partial y}$ e $\frac{\partial \mu^*}{\partial y}$ através da solução do sistema:

$$\begin{pmatrix} -\mu^* \cdot \frac{\partial^2 f^*}{\partial x_1^2} & -\mu^* \cdot \frac{\partial^2 f^*}{\partial x_2 \partial x_1} & -\frac{\partial f^*}{\partial x_1} \\ -\mu^* \cdot \frac{\partial^2 f^*}{\partial x_1 \partial x_2} & -\mu^* \cdot \frac{\partial^2 f^*}{\partial x_2^2} & -\frac{\partial f^*}{\partial x_2} \\ -\frac{\partial f^*}{\partial x_1} & -\frac{\partial f^*}{\partial x_2} & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial x_1^*}{\partial y} \\ \frac{\partial x_2^*}{\partial y} \\ \frac{\partial \mu^*}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

4.4 Teoria do consumidor: efeito-renda e efeito-substituição

Pelo Teorema de Arrow-Enthoven, se $|D_2| = \begin{vmatrix} 0 & u_1 & u_2 \\ u_1 & u_{11} & u_{12} \\ u_2 & u_{21} & u_{22} \end{vmatrix} > 0$, então u é estritamente quase-côncava. Suponhamos que $x_1(p_1, p_2, m) > 0$, $x_2(p_1, p_2, m) > 0$ e $\lambda(p_1, p_2, m) > 0$ satisfaçam as condições de Kuhn-Tucker para um máximo, ou seja, $p_1 \cdot x_1 + p_2 \cdot x_2 = m$, $u_1(x_1(p_1, p_2, m), x_2(p_1, p_2, m)) - \lambda(p_1, p_2, m) \cdot p_1 = 0$, $u_2(x_1(p_1, p_2, m), x_2(p_1, p_2, m)) - \lambda(p_1, p_2, m) \cdot p_2 = 0$.

Neste caso, observemos que $\begin{vmatrix} 0 & u_1 & u_2 \\ u_1 & u_{11} & u_{12} \\ u_2 & u_{21} & u_{22} \end{vmatrix} = \begin{vmatrix} 0 & -\lambda \cdot p_1 & -\lambda \cdot p_2 \\ -\lambda \cdot p_1 & u_{11} & u_{12} \\ -\lambda \cdot p_2 & u_{21} & u_{22} \end{vmatrix} = \lambda^2 \cdot \begin{vmatrix} 0 & p_1 & p_2 \\ p_1 & u_{11} & u_{12} \\ p_2 & u_{21} & u_{22} \end{vmatrix} > 0$. Se designarmos o determinante $\begin{vmatrix} 0 & p_1 & p_2 \\ p_1 & u_{11} & u_{12} \\ p_2 & u_{21} & u_{22} \end{vmatrix}$ por $|H_2|$, então $|D_2| = \lambda^2 \cdot |H_2| > 0$.

4.5 A curva de renda-consumo

Derivando, em relação à renda m , as condições acima, obtemos:

$$-p_1 \cdot \frac{\partial x_1}{\partial m} - p_2 \cdot \frac{\partial x_2}{\partial m} = -1,$$

$$u_{11}(x_1, x_2) \cdot \frac{\partial x_1}{\partial m} + u_{12}(x_1, x_2) \cdot \frac{\partial x_2}{\partial m} - \frac{\partial \lambda}{\partial m} \cdot p_1 = 0,$$

$$u_{21}(x_1, x_2) \cdot \frac{\partial x_1}{\partial m} + u_{22}(x_1, x_2) \cdot \frac{\partial x_2}{\partial m} - \frac{\partial \lambda}{\partial m} \cdot p_2 = 0.$$

Escrevendo na forma matricial, $\begin{pmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \lambda}{\partial m} \\ \frac{\partial x_1}{\partial m} \\ \frac{\partial x_2}{\partial m} \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$ e usando Cramer

obtemos $\frac{\partial x_1}{\partial m} = \frac{\begin{vmatrix} 0 & -1 & -p_2 \\ -p_1 & 0 & u_{12} \\ -p_2 & 0 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} = \frac{\begin{vmatrix} -p_1 & u_{12} \\ -p_2 & u_{22} \end{vmatrix}}{\begin{vmatrix} -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} = \frac{\begin{vmatrix} -p_1 & u_{12} \\ -p_2 & u_{22} \end{vmatrix}}{|H_2|}$. Dizemos que um bem é **normal** se

$\frac{\partial x_1}{\partial m} > 0$ e dizemos que um bem é **inferior** se $\frac{\partial x_1}{\partial m} < 0$.

Exercício 13

Obtenha $\frac{\partial x_2}{\partial m}$ e $\frac{\partial \lambda}{\partial m}$.

4.6 A curva de demanda-preço

Derivando em relação à renda p_1 , as condições de Kuhn-Tucker acima, obtemos:

$$x_1 + p_1 \cdot \frac{\partial x_1}{\partial p_1} + p_2 \cdot \frac{\partial x_2}{\partial p_1} = 0,$$

$$u_{11}(x_1, x_2) \cdot \frac{\partial x_1}{\partial p_1} + u_{12}(x_1, x_2) \cdot \frac{\partial x_2}{\partial p_1} - \frac{\partial \lambda}{\partial p_1} \cdot p_1 - \lambda = 0,$$

$$u_{21}(x_1, x_2) \cdot \frac{\partial x_1}{\partial p_1} + u_{22}(x_1, x_2) \cdot \frac{\partial x_2}{\partial p_1} - \frac{\partial \lambda}{\partial p_1} \cdot p_2 = 0,$$

Ou ainda, $\begin{pmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \lambda}{\partial p_1} \\ \frac{\partial x_1}{\partial p_1} \\ \frac{\partial x_2}{\partial p_1} \end{pmatrix} = \begin{pmatrix} x_1 \\ \lambda \\ 0 \end{pmatrix}$. Por Cramer

$$\frac{\partial x_1}{\partial p_1} = \frac{\begin{vmatrix} 0 & x_1 & -p_2 \\ -p_1 & \lambda & u_{12} \\ -p_2 & 0 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} = -x_1 \cdot \frac{\begin{vmatrix} -p_1 & u_{12} \\ -p_2 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} + \lambda \cdot \frac{\begin{vmatrix} 0 & -p_2 \\ -p_2 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}}$$

ou $\frac{\partial x_1}{\partial p_1} = \frac{\begin{vmatrix} 0 & x_1 & -p_2 \\ -p_1 & \lambda & u_{12} \\ -p_2 & 0 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} = -x_1 \cdot \frac{\partial x_1}{\partial m} + \lambda \cdot \frac{\begin{vmatrix} 0 & -p_2 \\ -p_2 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} = -x_1 \cdot \frac{\partial x_1}{\partial m} - \lambda \cdot \frac{p_2^2}{|H_2|}.$

A quantidade $-x_1 \cdot \frac{\partial x_1}{\partial m}$ chamamos “**efeito-renda**”. Observe que, se o bem é normal, o efeito-renda é uma quantidade negativa. Se o bem é inferior, o efeito-renda é positivo.

A quantidade:

$$-\lambda \cdot \frac{\begin{vmatrix} 0 & -p_2 \\ -p_2 & u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix}} = -\lambda \cdot \frac{p_2^2}{|H_2|}$$

chamamos “**efeito-substituição**”, e vamos explicá-la a seguir.

Digamos que fixemos o nível de utilidade em u^* e formulemos o seguinte problema:

$$\begin{aligned} \min \quad & p_1 \cdot h_1 + p_2 \cdot h_2 = E(p_1, p_2, u^*) \\ \text{sa} \quad & u(h_1, h_2) = u^* \\ & h_1 > 0, h_2 > 0. \end{aligned}$$

As condições de Kuhn-Tucker para este problema são:

$$u(h_1(p_1, p_2, u^*), h_2(p_1, p_2, u^*)) = u^*,$$

$$p_1 - \mu(p_1, p_2, u^*) \cdot u_1(h_1(p_1, p_2, u^*), h_2(p_1, p_2, u^*)) = 0,$$

$$p_2 - \mu(p_1, p_2, u^*) \cdot u_2(h_1(p_1, p_2, u^*), h_2(p_1, p_2, u^*)) = 0.$$

Derivando as condições acima em relação ao p_1 , obtemos:

$$u_1(h_1, h_2) \cdot \frac{\partial h_1}{\partial p_1} + u_2(h_1, h_2) \cdot \frac{\partial h_2}{\partial p_1} = 0,$$

$$1 - \frac{\partial \mu}{\partial p_1} \cdot u_1(h_1, h_2) - \mu \cdot u_{11}(h_1, h_2) \cdot \frac{\partial h_1}{\partial p_1} - \mu \cdot u_{12}(h_1, h_2) \cdot \frac{\partial h_2}{\partial p_1} = 0,$$

$$-\frac{\partial \mu}{\partial p_1} \cdot u_2(h_1, h_2) - \mu \cdot u_{21}(h_1, h_2) \cdot \frac{\partial h_1}{\partial p_1} - \mu \cdot u_{22}(h_1, h_2) \cdot \frac{\partial h_2}{\partial p_1} = 0.$$

Na forma matricial
$$\begin{pmatrix} 0 & u_1 & u_2 \\ u_1 & \mu \cdot u_{11} & \mu \cdot u_{12} \\ u_2 & \mu \cdot u_{21} & \mu \cdot u_{22} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \mu}{\partial p_1} \\ \frac{\partial h_1}{\partial p_1} \\ \frac{\partial h_2}{\partial p_1} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Calculando $\frac{\partial h_1}{\partial p_1} = \frac{\begin{vmatrix} 0 & 0 & u_2 \\ u_1 & 1 & \mu \cdot u_{12} \\ u_2 & 0 & \mu \cdot u_{22} \end{vmatrix}}{\begin{vmatrix} 0 & u_1 & u_2 \\ u_1 & \mu \cdot u_{11} & \mu \cdot u_{12} \\ u_2 & \mu \cdot u_{21} & \mu \cdot u_{22} \end{vmatrix}}.$ Observemos que $\mu = \frac{1}{\lambda}$ e, portanto,

$$\begin{vmatrix} 0 & u_1 & u_2 \\ u_1 & \mu \cdot u_{11} & \mu \cdot u_{12} \\ u_2 & \mu \cdot u_{21} & \mu \cdot u_{22} \end{vmatrix} = \begin{vmatrix} 0 & \frac{p_1}{\mu} & \frac{p_2}{\mu} \\ \frac{p_1}{\mu} & \mu \cdot u_{11} & \mu \cdot u_{12} \\ \frac{p_2}{\mu} & \mu \cdot u_{21} & \mu \cdot u_{22} \end{vmatrix} = \frac{1}{\mu} \cdot \begin{vmatrix} 0 & p_1 & p_2 \\ p_1 & u_{11} & u_{12} \\ p_2 & u_{21} & u_{22} \end{vmatrix} = \lambda \cdot |H_2|.$$

$$\frac{\partial h_1}{\partial p_1} = \frac{\begin{vmatrix} 0 & u_2 \\ u_2 & \mu \cdot u_{22} \end{vmatrix}}{\lambda \cdot |H_2|} = -\frac{\begin{vmatrix} 0 & -\lambda \cdot p_2 \\ -\lambda \cdot p_2 & \mu \cdot u_{22} \end{vmatrix}}{\lambda \cdot |H_2|} = -\lambda \cdot \frac{p_2^2}{|H_2|} < 0$$

Sendo assim, $\frac{\partial x_1}{\partial p_1} = -x_1 \cdot \frac{\partial x_1}{\partial m} + \frac{\partial h_1}{\partial p_1}$. Esse último resultado é conhecido como “Equação de Slutsky”¹⁸.

Exemplo 9: Podemos obter a Equação de Slutsky utilizando o Teorema do Envelope. Vejamos: combinando a solução do problema de maximização da utilidade com a solução do problema da minimização do dispêndio por parte do consumidor, obtemos $x_1^*(p_1, p_2, E(p_1, p_2, u^*)) = h_1(p_1, p_2, u^*)$. Derivando ambos os lados dessa última equação em relação ao p_1 , obtemos $\frac{\partial x_1^*}{\partial p_1} + \frac{\partial x_1^*}{\partial m} \cdot \frac{\partial E}{\partial p_1} = \frac{\partial h_1}{\partial p_1}$. Pelo Teorema do Envelope, $\frac{\partial E}{\partial p_1}(p_1, p_2, u^*) = h_1(p_1, p_2, u^*) = x_1^*(p_1, p_2, m)$. Logo, $\frac{\partial x_1^*}{\partial p_1} + \frac{\partial x_1^*}{\partial m} \cdot x_1^* = \frac{\partial h_1}{\partial p_1}$, que é a equação de Slutsky.

Exercício 14

Mostre que $\frac{\partial h_1}{\partial p_2} = \lambda \cdot \frac{p_1 \cdot p_2}{|H_2|} > 0$.

4.7 A metodologia primal-dual de Eugene Silberberg¹⁹ (caso com restrições)

Vamos usar o problema de minimização de custos para expor a metodologia de Silberberg para Estática Comparativa nos casos com restrições. Novamente, consideremos:

$$\begin{aligned} \min \quad & w_1 \cdot x_1 + w_2 \cdot x_2 = C(w_1, w_2, y) \\ \text{s.a.} \quad & f(x_1, x_2) \geq y \\ & x_1 > 0, x_2 > 0, \end{aligned}$$

e digamos que as demandas indiretas ótimas existam $x_1^*(w_1, w_2, y) > 0$, $x_2^*(w_1, w_2, y) > 0$ e que, associado a essas demandas, tenhamos $\mu^*(w_1, w_2, y) \geq 0$.

Silberberg introduziu um problema equivalente:

$$\begin{aligned} \max \quad & C(w_1, w_2, y) - (w_1 \cdot x_1 + w_2 \cdot x_2) \\ \text{s.a.} \quad & f(x_1, x_2) - y \geq 0 \\ & w_1 > 0, w_2 > 0, y > 0, x_1 > 0, x_2 > 0. \end{aligned}$$

Se as demandas ótimas existem para $w_1 > 0$, $w_2 > 0$ e $y > 0$, então, obviamente, o problema acima tem solução, ou seja, $(x_1^*, x_2^*, w_1, w_2, y)$ são máximos locais de $[C(w_1, w_2, y) - (w_1 \cdot x_1 + w_2 \cdot x_2)]$, dadas as restrições. O lagrangeano desse problema é $\mathcal{L} = C(w_1, w_2, y) - w_1 \cdot x_1 - w_2 \cdot x_2 - \theta \cdot (f(x_1, x_2) - y)$ e são cinco variáveis de escolha: $x_1 > 0$, $x_2 > 0$, $w_1 > 0$, $w_2 > 0$ e $y > 0$. As condições de primeira ordem são:

$$\frac{\partial \mathcal{L}}{\partial x_1} = -w_1 - \theta \cdot f_{x_1} = 0, \quad \frac{\partial \mathcal{L}}{\partial x_2} = -w_2 - \theta \cdot f_{x_2} = 0, \quad \frac{\partial \mathcal{L}}{\partial w_1} = C_{w_1} - x_1 = 0, \quad \frac{\partial \mathcal{L}}{\partial w_2} = C_{w_2} - x_2 = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = C_y + \theta = 0, \quad \frac{\partial \mathcal{L}}{\partial \theta} = -f + y = 0.$$

A derivada segunda do lagrangeano é:

$$\mathcal{L}'' = \begin{pmatrix} -\theta \cdot f_{x_1 x_1} & -\theta \cdot f_{x_1 x_2} & -1 & 0 & 0 & -f_{x_1} \\ -\theta \cdot f_{x_2 x_1} & -\theta \cdot f_{x_2 x_2} & 0 & -1 & 0 & -f_{x_2} \\ -1 & 0 & C_{w_1 w_1} & C_{w_1 w_2} & C_{w_1 y} & 0 \\ 0 & -1 & C_{w_2 w_1} & C_{w_2 w_2} & C_{w_2 y} & 0 \\ 0 & 0 & C_{y w_1} & C_{y w_2} & C_{yy} & 1 \\ -f_{x_1} & -f_{x_2} & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Inicialmente, supusemos condições suficientes, ou seja, $\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & 0 \end{vmatrix} = D_2 > 0$,

para garantir os resultados de estática comparativa.

Por sua vez, Silberberg supôs a existência das soluções e impôs, como condição necessária, que $\mathcal{L}''$ seja negativa-semidefinida no subespaço dos vetores v ortogonais ao vetor gradiente $(f_{x_1} \ f_{x_2} \ 0 \ 0 \ -1 \ 0)$ da restrição $f(x_1, x_2) - y$. Nesse caso, todos os menores principais orlados de ordem $p \geq 2$ de $\mathcal{L}''$ são tais que $(-1)^p \cdot |H_p''| \geq 0$.

Por exemplo, $(-1)^2 \cdot \begin{vmatrix} \theta \cdot f_{x_1 x_1} & \theta \cdot f_{x_1 x_2} & f_{x_1} \\ \theta \cdot f_{x_2 x_1} & \theta \cdot f_{x_2 x_2} & f_{x_2} \\ f_{x_1} & f_{x_2} & 0 \end{vmatrix} = \theta \cdot \begin{vmatrix} f_{x_1 x_1} & f_{x_1 x_2} & f_{x_1} \\ f_{x_2 x_1} & f_{x_2 x_2} & f_{x_2} \\ f_{x_1} & f_{x_2} & 0 \end{vmatrix} \geq 0$. Como $\theta \geq 0$, então $D_2 = \begin{vmatrix} f_{x_1 x_1} & f_{x_1 x_2} & f_{x_1} \\ f_{x_2 x_1} & f_{x_2 x_2} & f_{x_2} \\ f_{x_1} & f_{x_2} & 0 \end{vmatrix} \geq 0$, que é uma condição que obtemos pelo Teorema de Arrow-Enthoven se exigirmos $f(x_1, x_2)$ quase-côncava.

Outro exemplo: os menores principais de $\begin{pmatrix} C_{w_1 w_1} & C_{w_1 w_2} & C_{w_1 y} & 0 \\ C_{w_2 w_1} & C_{w_2 w_2} & C_{w_2 y} & 0 \\ C_{y w_1} & C_{y w_2} & C_{yy} & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ seguem a mesma regra. Os menores principais orlados de ordem par são maiores ou iguais a zero e os menores de ordem ímpar menores ou iguais a zero. O

$\begin{vmatrix} C_{w_1 w_1} & C_{w_1 w_2} & C_{w_1 y} & 0 \\ C_{w_2 w_1} & C_{w_2 w_2} & C_{w_2 y} & 0 \\ C_{y w_1} & C_{y w_2} & C_{yy} & 1 \\ 0 & 0 & 1 & 0 \end{vmatrix} = \begin{vmatrix} C_{w_1 w_1} & C_{w_1 w_2} & 0 \\ C_{w_2 w_1} & C_{w_2 w_2} & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0$. Os menores orlados de ordem dois são

$\begin{vmatrix} C_{w_2 w_2} & C_{w_2 y} & 0 \\ C_{y w_2} & C_{yy} & 1 \\ 0 & 1 & 0 \end{vmatrix} = -C_{w_2 w_2} \geq 0$, $\begin{vmatrix} C_{w_1 w_1} & C_{w_1 y} & 0 \\ C_{y w_1} & C_{yy} & 1 \\ 0 & 1 & 0 \end{vmatrix} = -C_{w_1 w_1} \geq 0$ e $\begin{vmatrix} C_{w_1 w_1} & C_{w_1 w_2} & 0 \\ C_{w_2 w_1} & C_{w_2 w_2} & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0$. Pelo Teorema do Envelope, sabemos que $\frac{\partial C}{\partial w_1} = x_1^*$, $\frac{\partial C}{\partial w_2} = x_2^*$, portanto, $\frac{\partial^2 C}{\partial w_1^2} = \frac{\partial x_1^*}{\partial w_1} \leq 0$ e $\frac{\partial^2 C}{\partial w_2^2} = \frac{\partial x_2^*}{\partial w_2} \leq 0$. Observando, ainda pelo

Envelope, que $\frac{\partial C}{\partial y} = \mu^* = \theta^*$, então $\begin{vmatrix} (x_1^*)_{w_1} & (x_1^*)_{w_2} & (x_1^*)_y & 0 \\ (x_2^*)_{w_1} & (x_2^*)_{w_2} & (x_2^*)_y & 0 \\ \theta_{w_1}^* & \theta_{w_2}^* & \theta_y^* & 1 \\ 0 & 0 & 1 & 0 \end{vmatrix} = \begin{vmatrix} (x_1^*)_{w_1} & (x_1^*)_y & 0 \\ \theta_{w_1}^* & \theta_y^* & 1 \\ 0 & 1 & 0 \end{vmatrix} = -(x_1^*)_{w_1} \geq 0$, ou seja, $(x_2^*)_{w_2} \leq 0$ e $(x_1^*)_{w_1} \leq 0$.

Se estamos tratando o problema da determinação das demandas hicksianas, então o resultado acima traduz-se como $(h_2^*)_{p_2} \leq 0$ e $(h_1^*)_{p_1} \leq 0$.

Observe que $\frac{\partial x_1^*}{\partial w_2}(w_1, w_2, y) = \frac{\partial^2 C}{\partial w_2 \partial w_1} = \frac{\partial^2 C}{\partial w_1 \partial w_2} = \frac{\partial x_2^*}{\partial w_1}(w_1, w_2, y)$, que é dita relação de reciprocidade. Se estamos tratando das demandas hicksianas, então

$$\frac{\partial h_1^*}{\partial p_2}(p_1, p_2, u_0) = \frac{\partial h_2^*}{\partial p_1}(p_1, p_2, u_0).$$

Dado o problema com restrições, por exemplo:

$$\begin{aligned} \max_{x_1, x_2} \quad & f(x_1, x_2, \alpha) = V(\alpha) \\ \text{s.a.} \quad & g(x_1, x_2, \alpha) = c, \end{aligned}$$

Silberberg o substitui pelo problema “primal-dual” abaixo, uma óbvia referência à Programação Linear, por

$$\begin{aligned} \min_{x_1, x_2, \alpha} \quad & V(\alpha) - f(x_1, x_2, \alpha) \\ \text{s.a.} \quad & g(x_1, x_2, \alpha) = 0. \end{aligned}$$

No caso do primeiro problema, estamos supondo que as soluções ótimas $x_1^*(\alpha)$, $x_2^*(\alpha)$ e $\lambda^*(\alpha)$ existam e que $\lambda^*(\alpha)$ é o correspondente multiplicador associado às soluções $x_1^*(\alpha)$ e $x_2^*(\alpha)$ do lagrangeano $\mathcal{L} = f(x_1, x_2, \alpha) - \lambda \cdot g(x_1, x_2, \alpha)$. As condições de primeira ordem são: $\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda \cdot \frac{\partial g}{\partial x_1} = 0$, $\frac{\partial \mathcal{L}}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda \cdot \frac{\partial g}{\partial x_2} = 0$.

O lagrangeano do “problema primal-dual” é $\mathcal{L} = V(\alpha) - f(x_1, x_2, \alpha) - \theta \cdot g(x_1, x_2, \alpha)$. Como $x_1^*(\alpha)$, $x_2^*(\alpha)$ e $\lambda^*(\alpha)$ são ótimos do problema de minimização acima, pelas

condições de primeira ordem: $\frac{\partial \mathcal{L}}{\partial x_1} = -\frac{\partial f}{\partial x_1} - \theta \cdot \frac{\partial g}{\partial x_1} = 0$, $\frac{\partial \mathcal{L}}{\partial x_2} = -\frac{\partial f}{\partial x_2} - \theta \cdot \frac{\partial g}{\partial x_2} = 0$. Portanto, $\theta^* = -\lambda^*(\alpha)$. A condição $\frac{\partial \mathcal{L}}{\partial \alpha} = V'(\alpha) - \frac{\partial f}{\partial \alpha} - \theta \cdot \frac{\partial g}{\partial \alpha} = 0$ é o Teorema do Envelope: $V'(\alpha) = \frac{\partial f}{\partial \alpha}(x_1^*(\alpha), x_2^*(\alpha), \alpha) - \lambda^*(\alpha) \cdot \frac{\partial g}{\partial \alpha}(x_1^*(\alpha), x_2^*(\alpha), \alpha)$.

Pelo que vimos na Lição Sete, como estamos supondo que (x_1^*, x_2^*, α) é um mínimo local, então $\mathcal{L}''$ é positiva-semidefinida no subespaço ortogonal ao gradiente $\nabla g(x_1^*, x_2^*, \alpha)$. Portanto, todos os menores principais orlados de ordem $m \geq 2$, pois temos uma restrição apenas, da matriz orlada

$$\mathcal{L}''(x_1^*, x_2^*, \alpha) = \begin{pmatrix} -f_{x_1 x_1} - \theta \cdot g_{x_1 x_1} & -f_{x_1 x_2} - \theta \cdot g_{x_1 x_2} & -f_{x_1 \alpha} - \theta \cdot g_{x_1 \alpha} & -g_{x_1} \\ -f_{x_2 x_1} - \theta \cdot g_{x_2 x_1} & -f_{x_2 x_2} - \theta \cdot g_{x_2 x_2} & -f_{x_2 \alpha} - \theta \cdot g_{x_2 \alpha} & -g_{x_2} \\ -f_{\alpha x_1} - \theta \cdot g_{\alpha x_1} & -f_{\alpha x_2} - \theta \cdot g_{\alpha x_2} & V_{\alpha\alpha} - f_{\alpha\alpha} - \theta \cdot g_{\alpha\alpha} & -g_{\alpha} \\ -g_{x_1} & -g_{x_2} & -g_{\alpha} & 0 \end{pmatrix} \text{ são menores ou iguais a zero: } (-1) \cdot |H_r^{\pi}| \geq 0.$$

Em geral, temos um vetor $\alpha = (\alpha_1, \dots, \alpha_m)$ de parâmetros exógenos e a submatriz

$$\begin{pmatrix} V_{\alpha\alpha} - f_{\alpha\alpha} - \theta \cdot g_{\alpha\alpha} & -g_{\alpha} \\ -g_{\alpha} & 0 \end{pmatrix} = \begin{pmatrix} V_{\alpha_1 \alpha_1} - f_{\alpha_1 \alpha_1} - \theta \cdot g_{\alpha_1 \alpha_1} & \dots & V_{\alpha_1 \alpha_m} - f_{\alpha_1 \alpha_m} - \theta \cdot g_{\alpha_1 \alpha_m} & -g_{\alpha_1} \\ \vdots & \ddots & \vdots & \vdots \\ V_{\alpha_m \alpha_1} - f_{\alpha_m \alpha_1} - \theta \cdot g_{\alpha_m \alpha_1} & \dots & V_{\alpha_m \alpha_m} - f_{\alpha_m \alpha_m} - \theta \cdot g_{\alpha_m \alpha_m} & -g_{\alpha_m} \\ -g_{\alpha_1} & \dots & -g_{\alpha_m} & 0 \end{pmatrix} \text{ é tal que os menores orlados de}$$

ordem $m \geq 2$ são maiores ou iguais a zero. Todas as relações possíveis entre essas derivadas são estabelecidas através do fato que $|H_m| \leq 0$.

Consideremos um problema com mais restrições:

$$\begin{array}{ll} \max_{x_1, x_2, \dots, x_n} & f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = V(\alpha_1, \dots, \alpha_m) \\ \text{s.a.} & g_1(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0, \\ & g_2(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0, \\ & \vdots \\ & g_k(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0, \end{array}$$

tal que as variáveis de escolha sejam $x_1, x_2, \dots, x_n$ e $\alpha = (\alpha_1, \dots, \alpha_m)$ é um vetor de parâmetros exógenos. Substituímos o problema acima pelo problema:

$$\begin{array}{ll} \min_x & V(\alpha) - f(x, \alpha) = V(\alpha) \\ \text{s.a.} & g_1(x, \alpha) = 0, \\ & g_2(x, \alpha) = 0, \\ & \vdots \\ & g_k(x, \alpha) = 0. \end{array}$$

No primeiro problema, estamos supondo que as soluções ótimas $x_1^*(\alpha)$, $x_2^*(\alpha)$, $\dots$, $x_n^*(\alpha)$ e que $\lambda_1^*(\alpha)$, $\lambda_2^*(\alpha)$, $\dots$, $\lambda_k^*(\alpha)$ são os correspondentes multiplicadores associados às soluções $x_1^*(\alpha)$, $x_2^*(\alpha)$, $\dots$, $x_n^*(\alpha)$ do lagrangeano $\mathcal{L} = f(x, \alpha) - \lambda_1 \cdot g_1(x, \alpha) - \dots - \lambda_k \cdot g_k(x, \alpha)$.

O lagrangeano do “problema primal-dual” é $\mathcal{L} = V(\alpha) - f(x, \alpha) - \theta_1 \cdot g_1(x, \alpha) - \dots - \theta_k \cdot g_k(x, \alpha)$, como $x^*(\alpha) = (x_1^*(\alpha), \dots, x_n^*(\alpha))$, $\lambda_1^*(\alpha), \dots, \lambda_k^*(\alpha)$ satisfazem o problema de maximização acima, pelas condições de primeira ordem: $-\frac{\partial f}{\partial x_1} - \theta_1 \cdot \frac{\partial g_1}{\partial x_1} - \dots - \theta_k \cdot \frac{\partial g_k}{\partial x_1} = 0$, $-\frac{\partial f}{\partial x_2} - \theta_1 \cdot \frac{\partial g_1}{\partial x_2} - \dots - \theta_k \cdot \frac{\partial g_k}{\partial x_2} = 0$, $\dots$, $-\frac{\partial f}{\partial x_n} - \theta_1 \cdot \frac{\partial g_1}{\partial x_n} - \dots - \theta_k \cdot \frac{\partial g_k}{\partial x_n} = 0$. Essas equações determinam, univocamente, os parâmetros $\theta_l, l=1, \dots, k$ e $\theta_l^* = -\lambda_l^*(\alpha), l=1, \dots, k$. As condições $\frac{\partial \mathcal{L}}{\partial \alpha_i} = \frac{\partial V}{\partial \alpha_i} - \frac{\partial f}{\partial \alpha_i} - \theta_1 \cdot \frac{\partial g_1}{\partial \alpha_i} - \dots - \theta_k \cdot \frac{\partial g_k}{\partial \alpha_i} = 0, i=1, \dots, m$ são o Teorema do Envelope. A exigência que o problema impõe é que $\mathcal{L}^*$ é positiva-semidefinida no subespaço ortogonal ao gradiente $\nabla g(x_1^*, x_2^*, \alpha)$. Portanto, todos os menores principais de ordem $k+1$ – pois temos k restrições – da

matriz orlada $\mathcal{L}^*(x^*, \alpha) = \begin{pmatrix} \mathcal{L}_{xx}^*(x^*, \alpha) & \mathcal{L}_{x\alpha}^*(x^*, \alpha) & (g_1)_x & \dots & (g_k)_x \\ \mathcal{L}_{\alpha x}^*(x^*, \alpha) & \mathcal{L}_{\alpha\alpha}^*(x^*, \alpha) & (g_1)_\alpha & \dots & (g_k)_\alpha \\ (g_1)_x & (g_1)_\alpha & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (g_k)_x & (g_k)_\alpha & 0 & \dots & 0 \end{pmatrix}$, são maiores ou iguais a zero.

Em particular, os menores principais de ordem $k+1$ da matriz $\mathcal{L}^*(x^*, \alpha) = \begin{pmatrix} V_{\alpha_1 \alpha_1} - f_{\alpha_1 \alpha_1} + \sum_{l=1}^k \lambda_l \cdot (g_l)_{\alpha_1 \alpha_1} & \dots & V_{\alpha_1 \alpha_m} - f_{\alpha_1 \alpha_m} + \sum_{l=1}^k \lambda_l \cdot (g_l)_{\alpha_1 \alpha_m} & -(g_1)_{\alpha_1} & \dots & -(g_k)_{\alpha_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ V_{\alpha_m \alpha_1} - f_{\alpha_m \alpha_1} + \sum_{l=1}^k \lambda_l \cdot (g_l)_{\alpha_m \alpha_1} & \dots & V_{\alpha_m \alpha_m} - f_{\alpha_m \alpha_m} + \sum_{l=1}^k \lambda_l \cdot (g_l)_{\alpha_m \alpha_m} & -(g_1)_{\alpha_m} & \dots & -(g_k)_{\alpha_m} \\ \hline -(g_1)_{\alpha_1} & \dots & -(g_1)_{\alpha_m} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -(g_k)_{\alpha_1} & \dots & -(g_k)_{\alpha_m} & 0 & \dots & 0 \end{pmatrix}$ são tais que têm o sinal $(-1)^k$ ou são nulos.

Exemplo 10: No caso da maximização do consumidor, a matriz acima traduz-se como: $\begin{pmatrix} u_{p_1 p_1}^* & \dots & u_{p_1 p_n}^* & u_{p_1 m}^* & -x_1^* \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ u_{p_n p_1}^* & \dots & u_{p_n p_n}^* & u_{p_n m}^* & -x_n^* \\ u_{mp_1}^* & \dots & u_{mp_n}^* & u_{mm}^* & 1 \\ -x_1^* & \dots & -x_n^* & 1 & 0 \end{pmatrix}$. Todos os menores principais orlados de ordem $k \geq 2$ são maiores ou iguais a zero.

No caso da minimização da função custo, a matriz traduz-se como: $\begin{pmatrix} C_{w_1 w_1} & \dots & C_{w_1 w_n} & C_{w_1 y} & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ C_{w_n w_1} & \dots & C_{w_n w_n} & C_{w_n y} & 0 \\ C_{y w_1} & \dots & C_{y w_n} & C_{yy} & 1 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$. Todos os menores principais orlados $|H_k|$ de ordem $k \geq 2$ são tais que $(-1)^k \cdot |H| \geq 0$.

4.8 Explicando Silberberg

Consideremos a metodologia de Silberberg com algum detalhe para quem tiver interesse na formulação primal-dual. Seja o problema:

$$\begin{aligned} \max_{x_1, \dots, x_n} \quad & f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = \phi(\alpha_1, \dots, \alpha_m) \\ \text{sa} \quad & g_1(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0 \\ & \vdots \\ & g_r(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0. \end{aligned}$$

Pelo Teorema do Envelope, $\frac{\partial \phi}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha), \alpha) - \lambda_1^*(\alpha) \cdot \frac{\partial g_1}{\partial \alpha_i}(x^*(\alpha), \alpha) - \dots - \lambda_r^*(\alpha) \cdot \frac{\partial g_r}{\partial \alpha_i}(x^*(\alpha), \alpha)$.

Derivando a expressão acima em relação à variável α_j , $\frac{\partial^2 \phi}{\partial \alpha_j \partial \alpha_i}(\alpha) =$

$$\begin{aligned} & \left(\sum_{k=1}^{k=r} \frac{\partial^2 f}{\partial x_k \partial \alpha_i}(x^*(\alpha), \alpha) \cdot \frac{\partial x_k^*}{\partial \alpha_j}(\alpha) \right) + \frac{\partial^2 f}{\partial \alpha_j \partial \alpha_i}(x^*(\alpha), \alpha) - \\ & \frac{\partial \lambda_1^*}{\partial \alpha_j}(\alpha) \cdot \frac{\partial g_1}{\partial \alpha_i}(x^*(\alpha), \alpha) - \lambda_1^*(\alpha) \cdot \left(\sum_{k=1}^{k=r} \frac{\partial^2 g_1}{\partial x_k \partial \alpha_i}(x^*(\alpha), \alpha) \cdot \frac{\partial x_k^*}{\partial \alpha_j}(\alpha) + \frac{\partial^2 g_1}{\partial \alpha_j \partial \alpha_i}(x^*(\alpha), \alpha) \right) - \\ & \frac{\partial \lambda_r^*}{\partial \alpha_j}(\alpha) \cdot \frac{\partial g_r}{\partial \alpha_i}(x^*(\alpha), \alpha) - \lambda_r^*(\alpha) \cdot \left(\sum_{k=1}^{k=r} \frac{\partial^2 g_r}{\partial x_k \partial \alpha_i}(x^*(\alpha), \alpha) \cdot \frac{\partial x_k^*}{\partial \alpha_j}(\alpha) + \frac{\partial^2 g_r}{\partial \alpha_j \partial \alpha_i}(x^*(\alpha), \alpha) \right) \end{aligned}$$

Em termos matriciais:

$$\begin{aligned} & \begin{pmatrix} \frac{\partial^2 \phi}{\partial \alpha_1 \partial \alpha_1} & \dots & \frac{\partial^2 \phi}{\partial \alpha_m \partial \alpha_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \phi}{\partial \alpha_1 \partial \alpha_m} & \dots & \frac{\partial^2 \phi}{\partial \alpha_m \partial \alpha_m} \end{pmatrix}_{m \times m} = \\ & \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial \alpha_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial \alpha_1} \\ \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial \alpha_m} & \dots & \frac{\partial^2 f}{\partial x_n \partial \alpha_m} \end{pmatrix}_{m \times n} \cdot \begin{pmatrix} \frac{\partial x_1^*}{\partial \alpha_1} & \dots & \frac{\partial x_n^*}{\partial \alpha_m} \\ \vdots & \dots & \vdots \\ \frac{\partial x_1^*}{\partial \alpha_1} & \dots & \frac{\partial x_n^*}{\partial \alpha_m} \end{pmatrix}_{n \times m} + \begin{pmatrix} \frac{\partial^2 f}{\partial \alpha_1 \partial \alpha_1} & \dots & \frac{\partial^2 f}{\partial \alpha_m \partial \alpha_1} \\ \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial \alpha_1 \partial \alpha_m} & \dots & \frac{\partial^2 f}{\partial \alpha_m \partial \alpha_m} \end{pmatrix}_{m \times m} - \\ & \begin{pmatrix} \frac{\partial g_1}{\partial \alpha_1} & \dots & \frac{\partial g_r}{\partial \alpha_1} \\ \vdots & \dots & \vdots \\ \frac{\partial g_1}{\partial \alpha_m} & \dots & \frac{\partial g_r}{\partial \alpha_m} \end{pmatrix}_{m \times r} \cdot \begin{pmatrix} \frac{\partial \lambda_1^*}{\partial \alpha_1} & \dots & \frac{\partial \lambda_1^*}{\partial \alpha_m} \\ \vdots & \dots & \vdots \\ \frac{\partial \lambda_r^*}{\partial \alpha_1} & \dots & \frac{\partial \lambda_r^*}{\partial \alpha_m} \end{pmatrix}_{r \times m} - \\ & \sum_{k=1}^{k=r} \lambda_k^* \cdot \begin{pmatrix} \frac{\partial^2 g_k}{\partial x_1 \partial \alpha_1} & \dots & \frac{\partial^2 g_k}{\partial x_n \partial \alpha_1} \\ \vdots & \dots & \vdots \\ \frac{\partial^2 g_k}{\partial x_1 \partial \alpha_m} & \dots & \frac{\partial^2 g_k}{\partial x_n \partial \alpha_m} \end{pmatrix}_{m \times n} \cdot \begin{pmatrix} \frac{\partial x_1^*}{\partial \alpha_1} & \dots & \frac{\partial x_n^*}{\partial \alpha_m} \\ \vdots & \dots & \vdots \\ \frac{\partial x_1^*}{\partial \alpha_1} & \dots & \frac{\partial x_n^*}{\partial \alpha_m} \end{pmatrix}_{n \times m} - \sum_{k=1}^{k=r} \lambda_k^* \cdot \begin{pmatrix} \frac{\partial^2 g_k}{\partial \alpha_1 \partial \alpha_1} & \dots & \frac{\partial^2 g_k}{\partial \alpha_m \partial \alpha_1} \\ \vdots & \dots & \vdots \\ \frac{\partial^2 g_k}{\partial \alpha_1 \partial \alpha_m} & \dots & \frac{\partial^2 g_k}{\partial \alpha_m \partial \alpha_m} \end{pmatrix}_{m \times m}. \end{aligned}$$

Sem dúvida, não é uma expressão simples. Silberberg a resume como $\phi_{aa} =$

$$\left(f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \cdots - \lambda_r^* \cdot (g_r)_{ax}^*\right) \cdot \left(\frac{\partial x^*}{\partial \alpha}\right) + f_{aa}^* - g_{\alpha}^* \cdot \frac{\partial \lambda^*}{\partial \alpha} - \lambda_1^* \cdot (g_1)_{aa}^* - \cdots - \lambda_r^* \cdot (g_r)_{aa}^*.$$

O lagrangeano do problema primal-dual

$$\begin{aligned} \min_{x_1, \dots, x_n, \alpha_1, \dots, \alpha_m} \quad & \phi(\alpha_1, \dots, \alpha_m) - f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) \\ \text{sa} \quad & g_1(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0 \\ & \vdots \\ & g_r(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m) = 0, \end{aligned}$$

é $\mathcal{L} = \phi(\alpha) - f(x, \alpha) - \sum_{k=1}^{k=r} \lambda_k \cdot g_k(x, \alpha)$, que é positiva-semidefinida no subespaço ortogonal aos vetores $\nabla g_k(x, \alpha)$. Portanto, dada a matriz:

$$H = \left(\begin{array}{ccc|ccc|ccc} \mathcal{L}_{x_1 x_1}^* & \cdots & \mathcal{L}_{x_1 x_n}^* & \mathcal{L}_{x_1 \alpha_1}^* & \cdots & \mathcal{L}_{x_1 \alpha_m}^* & -\frac{\partial g_1^*}{\partial x_1} & \cdots & -\frac{\partial g_r^*}{\partial x_1} \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \mathcal{L}_{x_n x_1}^* & \cdots & \mathcal{L}_{x_n x_n}^* & \mathcal{L}_{x_n \alpha_1}^* & \cdots & \mathcal{L}_{x_n \alpha_m}^* & -\frac{\partial g_1^*}{\partial x_n} & \cdots & -\frac{\partial g_r^*}{\partial x_n} \\ \hline \mathcal{L}_{\alpha_1 x_1}^* & \cdots & \mathcal{L}_{\alpha_1 x_n}^* & \mathcal{L}_{\alpha_1 \alpha_1}^* & \cdots & \mathcal{L}_{\alpha_1 \alpha_m}^* & -\frac{\partial g_1^*}{\partial \alpha_1} & \cdots & -\frac{\partial g_r^*}{\partial \alpha_1} \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \mathcal{L}_{\alpha_m x_1}^* & \cdots & \mathcal{L}_{\alpha_m x_n}^* & \mathcal{L}_{\alpha_m \alpha_1}^* & \cdots & \mathcal{L}_{\alpha_m \alpha_m}^* & -\frac{\partial g_1^*}{\partial \alpha_m} & \cdots & -\frac{\partial g_r^*}{\partial \alpha_m} \\ \hline -\frac{\partial g_1^*}{\partial x_1} & \cdots & -\frac{\partial g_1^*}{\partial x_n} & -\frac{\partial g_1^*}{\partial \alpha_1} & \cdots & -\frac{\partial g_1^*}{\partial \alpha_m} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial g_r^*}{\partial x_1} & \cdots & -\frac{\partial g_r^*}{\partial x_n} & -\frac{\partial g_r^*}{\partial \alpha_1} & \cdots & -\frac{\partial g_r^*}{\partial \alpha_m} & 0 & \cdots & 0 \end{array} \right),$$

$$\text{então } H_{aa} = \left(\begin{array}{ccc|cc} \mathcal{L}_{\alpha_1 \alpha_1}^* & \cdots & \mathcal{L}_{\alpha_1 \alpha_m}^* & -\frac{\partial g_1^*}{\partial \alpha_1} & \cdots & -\frac{\partial g_r^*}{\partial \alpha_1} \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \mathcal{L}_{\alpha_m \alpha_1}^* & \cdots & \mathcal{L}_{\alpha_m \alpha_m}^* & -\frac{\partial g_1^*}{\partial \alpha_m} & \cdots & -\frac{\partial g_r^*}{\partial \alpha_m} \\ \hline -\frac{\partial g_1^*}{\partial \alpha_1} & \cdots & -\frac{\partial g_1^*}{\partial \alpha_m} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial g_r^*}{\partial \alpha_1} & \cdots & -\frac{\partial g_r^*}{\partial \alpha_m} & 0 & \cdots & 0 \end{array} \right)$$

é tal que seus menores orlados de ordem

$r+1$ são tais que têm o sinal $(-1)^k$ ou são nulos. Obviamente, uma condição necessária é que $m \geq r+1$.

Notemos por $\mathcal{L}_{aa}^* = \begin{pmatrix} \mathcal{L}_{\alpha_1 \alpha_1}^* & \cdots & \mathcal{L}_{\alpha_1 \alpha_m}^* \\ \vdots & \ddots & \vdots \\ \mathcal{L}_{\alpha_m \alpha_1}^* & \cdots & \mathcal{L}_{\alpha_m \alpha_m}^* \end{pmatrix}$, como $\mathcal{L}_{aa}^* = \phi_{aa} - f_{aa}^* - \sum_{k=1}^r \lambda_k^* (g_k^*)_{aa}$ e $\phi_{aa} =$

$$\left(f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \cdots - \lambda_r^* \cdot (g_r)_{ax}^*\right) \cdot \left(\frac{\partial x^*}{\partial \alpha}\right) + f_{aa}^* - g_{\alpha}^* \cdot \frac{\partial \lambda^*}{\partial \alpha} - \lambda_1^* \cdot (g_1)_{aa}^* - \cdots - \lambda_r^* \cdot (g_r)_{aa}^*, \quad \text{então}$$

$\mathcal{L}_{aa}^* = \left(f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \cdots - \lambda_r^* \cdot (g_r)_{ax}^*\right) \cdot \left(\frac{\partial x^*}{\partial \alpha}\right) - g_{\alpha}^* \cdot \frac{\partial \lambda^*}{\partial \alpha}$. A matriz acima impõe todas as condições possíveis sobre as derivadas $\left(\frac{\partial x^*}{\partial \alpha}\right)$.

Exemplo 11: Suponhamos que $f(x, y, a)$ depende de três parâmetros, tais que x e y podem variar livremente e que a seja mantido fixo. Definimos:

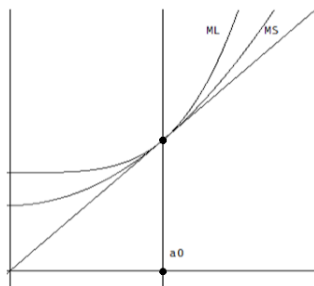
$$M_L(a) = \max_{x, y} f(x, y, a) \text{ e } M_S(y, a) = \max_x f(x, y, a),$$

ou seja, $M_L(a) = \max_{x, y} f(x, y, a)$ é um máximo obtido variando os dois parâmetros x e y ; $M_S(y, a) = \max_x f(x, y, a)$ variando apenas o parâmetro x com os parâmetros y e a fixados. Chamamos a função $M_L(a)$ de função valor-máximo de longo prazo e $M_S(y, a)$ de função valor-máximo de curto prazo. Obviamente, $M_L(a) \geq M_S(y, a)$. Fixado o parâmetro a em a_0 , sem dúvida, $M_L(a_0) = f(x^*(a_0), y^*(a_0), a_0) = M_S(y^*(a_0), a_0)$. Ou seja, fixado o valor do parâmetro y em $y^*(a_0)$, então o máximo de $f(x, y^*(a_0), a_0)$ está em $x^*(a_0)$, obviamente. Mas, para outros valores do parâmetro a , certamente,

$$M_L(a) = \max_{x, y} f(x, y, a) = f(x^*(a), y^*(a), a)$$

$$M_L(a) \geq M_S(y^*(a_0), a) = \max_x f(x, y^*(a_0), a) = f(x^*(y^*(a_0), a), y^*(a_0), a).$$

Abaixo, desenhamos um gráfico possível para $M_L(a)$ e $M_S(a)$:



Pelo Teorema do Envelope:

$$\frac{d}{da} M_L(a_0) = \frac{d}{da} f(x^*(a_0), y^*(a_0), a_0) = \frac{\partial}{\partial a} f(x^*(a_0), y^*(a_0), a_0).$$

Ou seja, em a_0 , as duas funções $M_L(a)$ e $M_S(y^*(a_0), a)$ têm a mesma derivada. Samuelson comenta esse fenômeno no capítulo 3 do livro “*Foundations of Economic Analysis*”, ao afirmar que “essa é a familiar relação de tangência entre o envelope de uma família de curvas e as curvas que ele toca”²⁰. Em outras palavras, no ponto em que a curva de curto prazo se ajusta à curva de longo prazo, as derivadas dessas duas funções nesse ponto, em relação ao

parâmetro exógeno, independem dos parâmetros que estão fixados na curva de curto prazo.

Como a função $[M_L(a) - M_S(y^*(a_0), a)]$ tem um mínimo global em a_0 e $\frac{d}{da}[M_L(a) - M_S(y^*(a_0), a)]_{a=a_0} = 0$, somos tentados a afirmar que $\frac{d}{da}M_L(a) \leq \frac{d}{da}M_S(y^*(a_0), a)$, para $a \leq a_0$, e $\frac{d}{da}M_L(a) \geq \frac{d}{da}M_S(y^*(a_0), a)$, para $a \geq a_0$. Intuitivamente, esse resultado significa que a função valor de longo prazo, que tem menos restrições que a função de curto prazo, se ajusta mais rapidamente. Entretanto, isso pode não ser verdadeiro no caso geral.

Exemplo 12: A função $f(x) = x^4 \cdot \left(\sin \frac{1}{x}\right)^2, x \neq 0, f(0) = 0$ é tal que $\lim_{x \rightarrow 0} f(x) = 0$, portanto, f é uma função contínua. f é derivável, pois $f'(0) = \lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0} x^3 \cdot \sin \frac{1}{x} = 0$, e se $x \neq 0$, $f'(x) = 4 \cdot x^3 \cdot \left(\sin \frac{1}{x}\right)^2 - 2 \cdot x^2 \cdot \left(\sin \frac{1}{x}\right) \cdot \left(\cos \frac{1}{x}\right)$, f' é contínua. Observe que $x = 0$ é um mínimo para f . No entanto, f' assume valores negativos e positivos em qualquer vizinhança de zero. Portanto, não podemos afirmar que $f'(x_0) \leq f'(0) = 0 \leq f'(x_1)$, para $x_0 < 0 < x_1$.

Contudo, se supusermos regularidades suficientes – por exemplo, que $\frac{d^2}{da^2}M_L(a)$ exista e $\frac{d^2}{da^2}[M_L(a) - M_S(y^*(a_0), a)] \geq 0$ numa vizinhança de a_0 –, o resultado será verdadeiro. Observe que $\frac{d^2}{da^2}[M_L(a) - M_S(y^*(a_0), a)] \geq 0$ numa vizinhança de a_0 é equivalente a afirmar que $[M_L(a) - M_S(y^*(a_0), a)]$ é convexa numa vizinhança de a_0 ou, ainda, que $\frac{d}{da}[M_L(a) - M_S(y^*(a_0), a)]$ é não decrescente na vizinhança de a_0 . Nesse caso, como $\frac{d}{da}[M_L(a) - M_S(y^*(a_0), a)]_{a=a_0} = 0$, então $\frac{d}{da}M_L(a) \leq \frac{d}{da}M_S(y^*(a_0), a)$ para $a \leq a_0$ e $\frac{d}{da}M_L(a) \geq \frac{d}{da}M_S(y^*(a_0), a)$ para $a \geq a_0$.

Digamos que $f: \mathbb{R} \rightarrow \mathbb{R}$ seja tal que $f(0) = f'(0) = 0, f''(x) \geq 0$ numa vizinhança de zero, então $f'(x) \leq 0$, para $x < 0$, e $f'(x) \geq 0$, para $x > 0$. Vejamos: $f'(x) = f'(0) + f''(\bar{x}) \cdot x = f''(\bar{x}) \cdot x$, para $\bar{x}$ entre 0 e x na proximidade de zero, então $f'(x) \leq 0$, para $x < 0$, e $f'(x) \geq 0$, para $x > 0$.

5. Princípio de Le Chatelier-Samuelson

Voltemos ao problema $V_L(\alpha_1, \dots, \alpha_m) = \max_{x_1, \dots, x_n} f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m)$, dito de longo prazo. Podemos considerar o problema de curto prazo $V_s(\alpha_1, \dots, \alpha_m) = \max_{x_0, \dots, x_n} f(x_1, \dots, x_n, \alpha_1, \dots, \alpha_m)$, em que as $(n-k)$ variáveis $x_l, l \neq i_1, \dots, i_k$ foram fixadas nos valores ótimos do problema

de longo prazo: se $\alpha = \alpha_0$ e $l \neq i_1, \dots, i_k$, então $x_l = x_l^*(\alpha_0)$, em que $x_l^*(\alpha_0)$ é o valor ótimo da variável x_l no problema de longo prazo.

Evidentemente, $V_s(\alpha_0) = V_L(\alpha_0)$, e para valores $\alpha \neq \alpha_0$, $V_s(\alpha) \leq V_L(\alpha)$. Pelo Teorema do Envelope, $\frac{\partial V_L}{\partial \alpha_r}(\alpha_0) = \frac{\partial f}{\partial \alpha_r}(x_1^*(\alpha_0), \dots, x_n^*(\alpha_0))$ e $\frac{\partial V_s}{\partial \alpha_r}(\alpha) = \frac{\partial f}{\partial \alpha_r}(\dots, x_{i_1}^*(\alpha), \dots, x_{i_k}^*(\alpha), \dots)$, sendo que as outras variáveis estão fixadas em $x_l = x_l^*(\alpha_0)$. Portanto, $\frac{\partial V_s}{\partial \alpha_r}(\alpha_0) = \frac{\partial f}{\partial \alpha_r}(\dots, x_{i_1}^*(\alpha_0), \dots, x_{i_k}^*(\alpha_0), \dots) = \frac{\partial V_L}{\partial \alpha_r}(\alpha_0)$.

Como $V_s(\alpha) - V_L(\alpha) \leq 0$ e $V_s(\alpha_0) = V_L(\alpha_0)$, então α_0 é um máximo global para $V_s(\alpha) - V_L(\alpha)$. Além disso, como $\frac{\partial V_s}{\partial \alpha_r}(\alpha_0) = \frac{\partial V_L}{\partial \alpha_r}(\alpha_0)$, então α_0 é um ponto crítico de $V_s(\alpha) - V_L(\alpha)$, então $(V_s - V_L)''(\alpha_0)$ é negativa-semidefinida. Repetiremos uma argumentação anterior: seja $\alpha_0 \in \square^m$ um ponto de máximo local para f , que é um ponto crítico de f ; seja uma trajetória $\varphi(t) = \alpha_0 + t \cdot v$, $v \in \square^m$, então $0 \leq \lim_{t \rightarrow 0} \frac{f(\varphi(t)) - f(\alpha_0)}{t^2} = \lim_{t \rightarrow 0} \frac{1}{2} \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right)' \cdot f''(\bar{\alpha}) \cdot \left(\frac{\varphi(t) - \varphi(0)}{t} \right) = \frac{1}{2} \cdot v' \cdot f''(\alpha_0) \cdot v$, ou seja, $v' \cdot f''(\alpha_0) \cdot v \leq 0$ é negativa-semidefinida. Logo, $(V_s - V_L)_{\alpha_i \alpha_j}(\alpha_0) \leq 0, i = 1, \dots, m$, pois essas derivadas são os menores principais de primeira ordem. Esse resultado é conhecido como “Princípio de Le Chatelier-Samuelson”. Quanto às demais derivadas $(V_s - V_L)_{\alpha_i \alpha_j}(\alpha_0), i \neq j$, que não estão na diagonal principal da matriz $(V_s - V_L)''(\alpha_0)$, não podemos fazer nenhuma afirmação; a única exceção é feita ao caso $(V_s - V_L)_{\alpha_i \alpha_i}(\alpha_0) = (V_s - V_L)_{\alpha_i \alpha_i}(\alpha_0)$. O comportamento da derivada $(V_s - V_L)''$ na vizinhança de α_0 depende de hipóteses adicionais como observamos no Exemplo 12.

5.1 Samuelson e um teorema “bem conhecido”

No capítulo III do “Análise”, seção “Restrições Auxiliares e o Princípio de Le Chatelier”, Samuelson analisa o que acontece com um equilíbrio determinado por um máximo, ou seja, por um ponto crítico de uma função côncava quando são acrescentadas restrições ao problema inicial que não alteram o equilíbrio. A razão desse interesse justifica-se “no fato de que nos permite deduzir as condições necessárias que o equilíbrio tem que satisfazer”²¹. Samuelson não faz referência, mas a utilidade que atribui ao seu procedimento é a mesma que em Mecânica se atribui aos chamados

“deslocamentos virtuais”, ou seja, obter informação sobre o “equilíbrio” sem que o deslocamento seja efetuado.

Samuelson faz suas considerações analisando um problema particular:

$$\phi(\alpha_1, \dots, \alpha_n) = \max_{x_1, \dots, x_n} f(x_1, \dots, x_n) - \alpha_1 \cdot x_1 - \alpha_2 \cdot x_2 - \dots - \alpha_n \cdot x_n.$$

As condições de primeira ordem desse problema são: $f_1 - \alpha_1 = 0$, $f_2 - \alpha_2 = 0$, ..., $f_n - \alpha_n = 0$. Suponhamos que existam as soluções $x^*(\alpha) = x^*(\alpha_1, \dots, \alpha_n) = (x_1^*(\alpha), \dots, x_n^*(\alpha))$ para o sistema de equações determinado pelas condições de primeira ordem, isto é,

$$\begin{aligned} f_1(x_1^*(\alpha), \dots, x_n^*(\alpha)) &= \alpha_1 \\ &\vdots \\ f_n(x_1^*(\alpha), \dots, x_n^*(\alpha)) &= \alpha_n. \end{aligned}$$

Pelo Teorema da Função Implícita, esse sistema tem solução, ou seja, é possível localmente expressar as variáveis $x = (x_1, \dots, x_n)$ como funções $\alpha = (\alpha_1, \dots, \alpha_n)$ se $\begin{vmatrix} f_{11} & \dots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \dots & f_{nn} \end{vmatrix} \neq 0$. Se supormos que f seja estritamente côncava, então o sistema tem uma única solução para cada $\alpha = (\alpha_1, \dots, \alpha_n)$ fixado.

Derivamos as condições de primeira ordem acima em relação à variável α_j para obter $\frac{\partial}{\partial \alpha_j} x^*(\alpha)$:

$$\begin{aligned} f_{11} \cdot \frac{\partial x_1^*}{\partial \alpha_j} + \dots + f_{1j} \cdot \frac{\partial x_j^*}{\partial \alpha_j} + \dots + f_{1n} \cdot \frac{\partial x_n^*}{\partial \alpha_j} &= 0 \\ &\vdots \\ f_{j1} \cdot \frac{\partial x_1^*}{\partial \alpha_j} + \dots + f_{jj} \cdot \frac{\partial x_j^*}{\partial \alpha_j} + \dots + f_{jn} \cdot \frac{\partial x_n^*}{\partial \alpha_j} &= 1 \\ &\vdots \\ f_{n1} \cdot \frac{\partial x_1^*}{\partial \alpha_j} + \dots + f_{nj} \cdot \frac{\partial x_j^*}{\partial \alpha_j} + \dots + f_{nn} \cdot \frac{\partial x_n^*}{\partial \alpha_j} &= 0, \end{aligned}$$

isto é, $\begin{pmatrix} f_{11}^* & \dots & f_{1n}^* \\ \vdots & \ddots & \vdots \\ f_{n1}^* & \dots & f_{nn}^* \end{pmatrix} \begin{pmatrix} \vdots \\ \frac{\partial x_j^*}{\partial \alpha_j} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ 1 \\ \vdots \end{pmatrix} \leftarrow j$, em que $f_{ij}^* = f_{ij}(x^*)$. Observe-se que Samuelson não

está analisando o comportamento de $\phi(\alpha_1, \dots, \alpha_n)$, mas de $x^*(\alpha)$. Por Cramer,

$\frac{\partial x_j^*}{\partial \alpha_j} = \frac{\Delta_{jj}}{\begin{vmatrix} f_{11}^* & \dots & f_{1n}^* \\ \vdots & \ddots & \vdots \\ f_{n1}^* & \dots & f_{nn}^* \end{vmatrix}} = \frac{|f''(j|j)|}{|f''|}$, em que $|f''(j|j)|$ é o determinante menor principal de

$f^{**} = \begin{pmatrix} f_{11}^* & \cdots & f_{1n}^* \\ \vdots & \ddots & \vdots \\ f_{n1}^* & \cdots & f_{nn}^* \end{pmatrix}$ obtido pela retirada da linha j e da coluna j . Como f é estritamente côncava, então $|f^{**}|$ e $|f^{**}(j|j)|$ têm sinais opostos. Portanto, $\frac{\partial x_j^*}{\partial \alpha_j} < 0$.

Exercício 15: Mostre que se $i \neq j$, então $\frac{\partial x_i^*}{\partial \alpha_j} = \frac{(-1)^{j+i} \cdot |f^{**}(j|i)|}{|f^{**}|} = \frac{\Delta_{ji}}{|f^{**}|}$, em que $(-1)^{j+i} \cdot |f^{**}(j|i)| = \Delta_{ji}$ é o cofator de f_{ji}^* na matriz f^{**} . Observe que $|f^{**}(j|j)| = \Delta_{jj}$. Por exemplo, observe que se $x = (x_1, x_2)$, então

$$\begin{aligned} \frac{\partial x_1^*}{\partial \alpha_1} &= \frac{\Delta_{11}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{\begin{vmatrix} 1 & f_{12}^* \\ 0 & f_{22}^* \end{vmatrix}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{f_{22}^*}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} < 0, & \frac{\partial x_1^*}{\partial \alpha_2} &= \frac{\Delta_{21}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{\begin{vmatrix} 0 & f_{12}^* \\ 1 & f_{22}^* \end{vmatrix}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{-f_{12}^*}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}}, & \frac{\partial x_2^*}{\partial \alpha_1} &= \frac{\Delta_{12}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{\begin{vmatrix} f_{11}^* & 1 \\ f_{21}^* & 0 \end{vmatrix}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{-f_{21}^*}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}}, \\ \frac{\partial x_2^*}{\partial \alpha_2} &= \frac{\Delta_{22}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{\begin{vmatrix} f_{11}^* & 0 \\ f_{21}^* & 1 \end{vmatrix}}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} = \frac{f_{11}^*}{\begin{vmatrix} f_{11}^* & f_{12}^* \\ f_{21}^* & f_{22}^* \end{vmatrix}} < 0. \end{aligned}$$

Em relação aos sinais de $\frac{\partial x_1^*}{\partial \alpha_2}$ e $\frac{\partial x_2^*}{\partial \alpha_1}$, não podemos afirmar nada *a priori*, pois, por exemplo, $f(x_1, x_2) = -x_1^2 + x_1 \cdot x_2 - x_2^2$, $f(x_1, x_2) = -x_1^2 - x_1 \cdot x_2 - x_2^2$ e $f(x_1, x_2) = -x_1^2 - x_2^2$ são estritamente convexas e as derivadas acima são respectivamente, negativas, positivas e nulas.

Suponhamos que acrescentemos ao problema inicial $\max_{x_1, \dots, x_n} f(x_1, \dots, x_n) - \alpha_1 \cdot x_1 - \alpha_2 \cdot x_2 - \dots - \alpha_n \cdot x_n$ uma restrição $g(x_1, \dots, x_n) = 0$, de maneira que a solução desse novo problema seja a mesma do problema inicial $x^*(\alpha_0) = (x_1^*(\alpha_0), \dots, x_n^*(\alpha_0))$, em que $\alpha = \alpha_0$, ou seja, deixamos o equilíbrio inalterado.

As condições de primeira ordem desse novo problema são: $f_1 - \lambda \cdot g_1 - \alpha_1 = 0$, $f_2 - \lambda \cdot g_2 - \alpha_2 = 0, \dots, f_n - \lambda \cdot g_n - \alpha_n = 0$, em que $\mathcal{L} = f - \alpha_1 \cdot x_1 - \dots - \alpha_n \cdot x_n - \lambda \cdot g$ é o lagrangeano do problema.

Para obter as derivadas das soluções $x^*(\alpha) = (x_1^*(\alpha), \dots, x_n^*(\alpha))$, em que $x^*(\alpha)$ é a solução do problema com a restrição, estabelecemos o sistema

$$\begin{aligned} (f_{11} - \lambda \cdot g_{11}) \cdot \frac{\partial^1 x_1^*}{\partial \alpha_j} + \dots + (f_{1j} - \lambda \cdot g_{1j}) \cdot \frac{\partial^1 x_j^*}{\partial \alpha_j} + \dots + (f_{1n} - \lambda \cdot g_{1n}) \cdot \frac{\partial^1 x_n^*}{\partial \alpha_j} - \frac{\partial \lambda}{\partial \alpha_j} \cdot g_1 &= 0 \\ \vdots & \\ (f_{j1} - \lambda \cdot g_{j1}) \cdot \frac{\partial^1 x_1^*}{\partial \alpha_j} + \dots + (f_{jj} - \lambda \cdot g_{jj}) \cdot \frac{\partial^1 x_j^*}{\partial \alpha_j} + \dots + (f_{jn} - \lambda \cdot g_{jn}) \cdot \frac{\partial^1 x_n^*}{\partial \alpha_j} - \frac{\partial \lambda}{\partial \alpha_j} \cdot g_j &= 1 \\ \vdots & \\ (f_{n1} - \lambda \cdot g_{n1}) \cdot \frac{\partial^1 x_1^*}{\partial \alpha_j} + \dots + (f_{nj} - \lambda \cdot g_{nj}) \cdot \frac{\partial^1 x_j^*}{\partial \alpha_j} + \dots + (f_{nn} - \lambda \cdot g_{nn}) \cdot \frac{\partial^1 x_n^*}{\partial \alpha_j} - \frac{\partial \lambda}{\partial \alpha_j} \cdot g_n &= 0 \\ (-g_1) \cdot \frac{\partial^1 x_1^*}{\partial \alpha_j} + \dots + (-g_j) \cdot \frac{\partial^1 x_j^*}{\partial \alpha_j} + \dots + (-g_n) \cdot \frac{\partial^1 x_n^*}{\partial \alpha_j} &= 0, \end{aligned}$$

isto é,

$$\begin{pmatrix} \mathcal{L}_{11}^* & \cdots & \mathcal{L}_{1n}^* & -g_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{L}_{n1}^* & \cdots & \mathcal{L}_{nn}^* & -g_n \\ -g_1 & \cdots & -g_n & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial^1 x_j^*}{\partial \alpha_j} \\ \vdots \\ \frac{\partial \lambda^*}{\partial \alpha_j} \end{pmatrix} = \begin{pmatrix} \vdots \\ 1 \\ \vdots \end{pmatrix} \leftarrow j, \quad ,$$

portanto, $\frac{\partial^1 x_j^*}{\partial \alpha_j} = \frac{\left| \begin{pmatrix} \mathcal{L}_{11}^* & \cdots & \mathcal{L}_{1n}^* & -g_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{L}_{n1}^* & \cdots & \mathcal{L}_{nn}^* & -g_n \\ -g_1 & \cdots & -g_n & 0 \end{pmatrix} (j, j) \right|}{\left| \begin{pmatrix} \mathcal{L}_{11}^* & \cdots & \mathcal{L}_{1n}^* & -g_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{L}_{n1}^* & \cdots & \mathcal{L}_{nn}^* & -g_n \\ -g_1 & \cdots & -g_n & 0 \end{pmatrix} \right|}$, em que $\left| \begin{pmatrix} \mathcal{L}_{11}^* & \cdots & \mathcal{L}_{1n}^* & -g_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{L}_{n1}^* & \cdots & \mathcal{L}_{nn}^* & -g_n \\ -g_1 & \cdots & -g_n & 0 \end{pmatrix} (j, j) \right|$ é o determinante menor principal de $\begin{pmatrix} \mathcal{L}_{11}^* & \cdots & \mathcal{L}_{1n}^* & -g_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{L}_{n1}^* & \cdots & \mathcal{L}_{nn}^* & -g_n \\ -g_1 & \cdots & -g_n & 0 \end{pmatrix} = |H|$ obtido pela eliminação da linha j e da coluna j , que é o cofator (j, j) , ${}^1\Delta_{jj}$, da matriz H .

Como o equilíbrio permaneceu inalterado pelo acréscimo da restrição, então, em relação às condições de primeira ordem do primeiro e do segundo problema, devemos ter que $f_i - \alpha_i = f_i - \lambda \cdot g_i - \alpha_i = 0$, ou seja, $\lambda^* = 0$, portanto,

$$\begin{pmatrix} f_{11}^* & \cdots & f_{1n}^* & -g_1 \\ \vdots & \ddots & \vdots & \vdots \\ f_{n1}^* & \cdots & f_{nn}^* & -g_n \\ -g_1 & \cdots & -g_n & 0 \end{pmatrix} = |H|.$$

$$\text{Decontando, } \frac{\partial^1 x_j^*}{\partial \alpha_j} - \frac{\partial x_j^*}{\partial \alpha_j} = \frac{|H(j|j)|}{|H|} - \frac{|f^{**}(j|j)|}{|f^{**}|} = \frac{|f^{**}| \cdot |H(j|j)| - |f^{**}(j|j)| \cdot |H|}{|f^{**}| \cdot |H|}.$$

Consideremos o caso particular em que $x = (x_1, x_2)$, então

$$\frac{\partial^1 x_1^*}{\partial \alpha_1} - \frac{\partial x_1^*}{\partial \alpha_1} = \frac{|H(1|1)|}{|H|} - \frac{|f^{**}(1|1)|}{|f^{**}|} = \frac{-(g_2^*)^2}{|H|} - \frac{f_{22}^*}{|f^{**}|} = \frac{-(g_2^*)^2 \cdot |f^{**}| - f_{22}^* \cdot |H|}{|H| \cdot |f^{**}|} \text{ e } |H| = \begin{vmatrix} f_{11}^* & f_{12}^* & -g_1^* \\ f_{21}^* & f_{22}^* & -g_2^* \\ -g_1^* & -g_2^* & 0 \end{vmatrix} |H| = \begin{vmatrix} f_{22}^* & -g_2^* \\ -g_2^* & f_{22}^* \end{vmatrix} = -(g_2^*)^2.$$

Nesse ponto, Samuelson emprega um teorema “bem conhecido”, devido a Carl Jacobi (1804-1851), publicado em 1833, que afirma que dada uma matriz qualquer $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$, então $\begin{vmatrix} \Delta_{hh} & \cdots & \Delta_{hk} \\ \vdots & \ddots & \vdots \\ \Delta_{hk} & \cdots & \Delta_{kk} \end{vmatrix} = |A|^{k-1} \cdot |A(i_1, i_2, \dots, i_k)|$, em que $\Delta_{ij} = (-1)^{i+j} \cdot |A(i|j)|$ é o cofator (i, j) da matriz A , $|A(i_1, i_2, \dots, i_k)|$ é menor obtido eliminando as linhas e colunas $i_1, i_2, \dots, i_k$ da matriz A e $|A|$ é o determinante de A . Esse teorema é uma generalização dos cofatores usuais da matriz A e é um resultado atualmente

desconhecido. Podemos encontrar esse resultado nos livros de Mirsky²² ou de Aitken²³, que são referências para outros autores, por exemplo, Eugene Silberberg²⁴, George Lady e James Quirk²⁵. Samuelson se refere a esse resultado como “um bem conhecido teorema sobre determinantes [de] Jacobi”²⁶. Samuelson também emprega esse resultado num artigo de 1960²⁷.

Voltando à expressão $\frac{\partial^1 x_1^*}{\partial \alpha_1} - \frac{\partial x_1^*}{\partial \alpha_1} = \frac{|H(1|1)|}{|H|} - \frac{|f''(1|1)|}{|f''|} = -\frac{(g_2^*)^2}{|H|} - \frac{f_{22}^*}{|f''|} = -\frac{(g_2^*)^2 \cdot |f''| - f_{22}^* \cdot |H|}{|H| \cdot |f''|}$, temos $\begin{vmatrix} \Delta_{11} & \Delta_{13} \\ \Delta_{31} & \Delta_{33} \end{vmatrix} = |H| \cdot f_{22}, \Delta_{11} = \begin{vmatrix} f_{22} & -g_2 \\ -g_2 & 0 \end{vmatrix} = -(g_2^*)^2, \Delta_{13} = \begin{vmatrix} f_{21} & f_{22} \\ -g_1 & -g_2 \end{vmatrix} = -g_2 \cdot f_{21} + g_1 \cdot f_{22}, \Delta_{31} = \begin{vmatrix} f_{12} & -g_1 \\ f_{22} & -g_2 \end{vmatrix} = -g_2 \cdot f_{12} + g_1 \cdot f_{22}$ e . Como , então . Logo, , então . A partir daí, concluímos que , pois , pois e . , ou .

A análise de Samuelson, que ilustramos com o caso acima, conclui que, se acrescentamos restrições sucessivamente , $^2g(x_1, \dots, x_n) = 0, \dots, ^{n-1}g(x_1, \dots, x_n) = 0$ ao problema inicial, obtemos . Samuelson observa: “isto explica porque as demandas de longo prazo são mais elásticas do que as a curto prazo”²⁸.

5.2 Os artigos de Silberberg e o princípio de Le Chatelier-Samuelson

Silberberg empregou a argumentação de Samuelson, ou seja, o Teorema de Jacobi, para analisar o comportamento das funções-valor $\phi_j(\alpha)$ no problema:

$$\begin{aligned} \max_{x_1, \dots, x_n} \quad & f(x_1, \dots, x_n, \alpha) = \phi^j(\alpha) \\ \text{sa} \quad & g_1(x_1, \dots, x_n) = 0 \\ & \vdots \\ & g_j(x_1, \dots, x_n) = 0, \end{aligned}$$

no qual as restrições são acrescentadas sucessivamente, $j=1, \dots, r$, a partir do problema inicial $\max_{x_1, \dots, x_n} f(x_1, \dots, x_n, \alpha) = \phi(\alpha)$, sem alterar a solução ótima $x^*(\alpha) = {}^1x^*(\alpha) \cdots = {}^{r-1}x^*(\alpha) = {}^rx^*(\alpha)$. O Princípio de Le Chatelier-Samuelson $\frac{\partial x_j^*}{\partial \alpha_j} \leq \frac{\partial^1 x_j^*}{\partial \alpha_j} \leq \frac{\partial^2 x_j^*}{\partial \alpha_j} \leq \dots \leq \frac{\partial^r x_j^*}{\partial \alpha_j} \leq 0$ é consequência, em situações particulares, do Teorema do

Envelope Generalizado, de Silberberg, sobre as funções-valor: $\phi^r(\alpha) = \phi^{r-1}(\alpha)$, e $\phi_{aa}^r(\alpha) < \phi_{aa}^{r-1}(\alpha)$ ²⁹.

Os resultados $\phi^r(\alpha) = \phi^{r-1}(\alpha)$ e são obtidos imediatamente. Por hipótese, a solução ótima é a mesma em todos os problemas e pelo Teorema do Envelope $\phi_a^j(\alpha) = f_a(x^*(\alpha), \alpha) - \lambda_1^* \cdot (g_1)_a(x^*(\alpha)) - \dots - \lambda_r^* \cdot (g_r)_a(x^*(\alpha))$, como não depende do parâmetro α , então $\phi_a^j(\alpha) = f_a(x^*(\alpha), \alpha)$.

Para mostrar que $\phi_{aa}^r(\alpha) < \phi_{aa}^{r-1}(\alpha)$, que é a parte difícil do resultado, Silberberg usa o Teorema de Jacobi. Silberberg postula que $f(x_1, \dots, x_n, \alpha)$ seja estritamente côncava nas variáveis $x_1, \dots, x_n$.

A argumentação de Silberberg não viabiliza a presença das variáveis exógenas nas restrições e, por essa razão, ele procurou superar essa dificuldade em outro artigo³⁰ utilizando o método primal-dual.

Consideremos o problema com restrições:

e o problema

que tem uma restrição adicional em relação ao primeiro.

Digamos que em $\alpha = \alpha_0$, a solução do primeiro problema é a mesma solução do segundo: , ou seja, a restrição não altera a solução em $\alpha = \alpha_0$. É claro que , para $\alpha \neq \alpha_0$, ou seja, tem um máximo global em $\alpha = \alpha_0$. Novamente, . Pelo Teorema do Envelope, como , , pois , então , .

Os lagrangeanos e dos problemas “dual-primal” acima são tais que em qualquer que satisfaz as restrições, pois .

Através da expressão $(f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \dots - \lambda_r^* \cdot (g_r)_{ax}^*) \cdot \left(\frac{\partial x^*}{\partial \alpha}\right) + f_{aa}^* - g_a^* \cdot \frac{\partial \lambda^*}{\partial \alpha} - \lambda_1^* \cdot (g_1)_{aa}^* - \dots - \lambda_r^* \cdot (g_r)_{aa}^*$, que obtivemos na seção 4.8 deste presente capítulo, então $L_{aa}^r = (f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \dots - \lambda_r^* \cdot (g_r)_{ax}^*) \cdot \left(\frac{\partial x^*}{\partial \alpha}\right) - g_a^* \cdot \frac{\partial \lambda^*}{\partial \alpha}$. Portanto, ${}^{r+1}L_{aa}^* - {}^rL_{aa}^* = ({}^{r+1}L_{aa}^* - {}^rL_{aa}^*)(x^*(\alpha_0), \alpha_0) = \phi_{aa}^{r+1}(\alpha_0) - \phi_{aa}^r(\alpha_0)$ é uma

matriz negativa-semidefinida, pois $(\phi_{r+1} - \phi_r)$ tem um máximo local em $\alpha = \alpha_0$. Observe que ${}^{r+1}L_{aa}^* = (f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \dots - \lambda_r^* \cdot (g_r)_{ax}^*) \cdot \left(\frac{\partial {}^{r+1}x^*}{\partial \alpha} \right) - {}^{r+1}g_a^* \cdot \frac{\partial \lambda^*}{\partial \alpha}$ e ${}^rL_{aa}^* = (f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \dots - \lambda_r^* \cdot (g_r)_{ax}^*) \cdot \left(\frac{\partial {}^rx^*}{\partial \alpha} \right) - {}^rg_a^* \cdot \frac{\partial \lambda^*}{\partial \alpha}$, em que os argumentos nas funções são $({}^{r+1}x^*(\alpha_0), \alpha_0) = ({}^rx^*(\alpha_0), \alpha_0) = (x^*(\alpha_0), \alpha_0)$. $(f_{ax}^* - \lambda_1^* \cdot (g_1)_{ax}^* - \dots - \lambda_r^* \cdot (g_r)_{ax}^*)$ é uma matriz $m \times n$. As matrizes $\left(\frac{\partial {}^{r+1}x^*}{\partial \alpha} \right)$ e $\left(\frac{\partial {}^rx^*}{\partial \alpha} \right)$ são ambas, mas não são formalmente iguais, pois presumivelmente, . Finalmente observe que e .

Um elemento elemento j na diagonal principal da é . Portanto, as inequações anteriores relacionam as derivadas $\frac{\partial {}^{r+1}x_k^*}{\partial \alpha_j}(\alpha_0)$ e .

Para obter o resultado $\phi_{aa}^r(\alpha) < \phi_{aa}^{r-1}(\alpha)$ é indispensável alguma hipótese adicional que no artigo anterior correspondeu à exigência que seja estritamente côncava nas variáveis $x_1, \dots, x_n$. No presente trabalho, a hipótese adicional tem de implicar que seja negativa-definida numa vizinhança de α_0 .

Finalizando

Este livro apresenta os conteúdos matemáticos básicos de Economia Matemática. Entretanto, atendemos este objetivo, digamos, “utilitário”, mostrando como a Teoria Econômica favoreceu o desenvolvimento da Matemática. A Programação Linear e Não-Linear, a Teoria dos Jogos, a análise das condições de segunda ordem, a análise da quase-convexidade e, finalmente, a Estática Comparativa impulsionaram a Matemática. Contudo, evidentemente, o assunto não está concluído. Abaixo, sugerimos alguns temas que estão na continuidade do presente livro.

No artigo³¹ de 1937, von Neumann empregou o Teorema do Teorema de Ponto Fixo de Brouwer para mostrar a existência de soluções para o modelo de crescimento nomeado “Modelo de von Neumann”. Este trabalho permitiu, através de uma generalização do Teorema de Brouwer, conhecida como “Teorema do Ponto Fixo de Kakutani”, uma nova demonstração para existência de selas em jogos de soma-zero³², a demonstração, devida a John F. Nash (1928-2015), da existência de equilíbrio em jogos não-

cooperativos³³ e, em 1954, a demonstração de existência de equilíbrio walrasiano por Lionel McKenzie³⁴, Arrow e Debreu³⁵. A análise da existência de equilíbrios em jogos e em sistemas walrasianos é uma sequência natural para o presente texto.

Os resultados matemáticos necessários para tratar os modelos de Vassili Leontief³⁶³⁷ (1905-1999) e de Piero Sraffa³⁸ (1898-1983) foram obtidos no começo do século vinte pelos matemáticos Ferdinand Georg Frobenius (1849-1917) e Oskar Perron (1880-1975). Estes resultados se referem ao que hoje chamamos Teoria de Matrizes Não-Negativas e passaram a serem empregados na Teoria Econômica na década de 1940. Primeiramente, David Hawkins (1913-2002) e Herbert Simon (1916-2001) publicaram o artigo “Note: Some Conditions of Macroeconomic Stability”³⁹, no qual os autores apresentam uma célebre condição para que a inversa de uma matriz seja não-negativa. A condição de Hawkins-Simon é a maneira mais acessível para obter os resultados fundamentais do Modelo de Leontief e de Sraffa. Pouco depois, inaugurou-se uma abordagem mais sofisticada para obter os mesmos resultados via Teorema de Ponto Fixo. O trabalho mais conhecido desta nova abordagem é o artigo “Nonnegative Square Matrices”⁴⁰ de Gerard Debreu de 1953.

Michio Morishima⁴¹ (1923-2004) utilizou a Dualidade em Programação Linear, o Modelo de von Neumann e o Teorema de Perron-Frobenius para apresentar uma versão matemática da Teoria do Valor-Trabalho e do Problema da Transformação na Teoria Econômica Marxista.

Um sequência para o presente livro é a chamada “Teoria da Produção”⁴² desenvolvida na escola neo-ricardiana.

Como enfatizamos nas notas históricas da Lição Sete, a análise das condições de segunda ordem estão relacionadas aos estudos empíricos sobre as demandas de commodities agrícolas por Henry L. Moore (1869-1958) e Henry Schultz⁴³ (1893-1938). Por sua vez, a análise da existência de

equilíbrio walrasiano relaciona-se ao Método Econométrico de Equações Simultâneas desenvolvido na Fundação Cowles.

Também, a Dinâmica Econômica está ligada ao nascimento da Econometria através das tentativas de superar os obstáculos para a análise empírica apresentadas pelas teorias estáticas. Além das dificuldades relacionadas à Teoria da Demanda somam-se as preocupações com a análise dos ciclos econômicos. Este é o objetivo dos trabalhos de Griffith Evans (1887-1973) nas aplicações do Cálculo Variacional à Teoria Econômica⁴⁴. Na década de 1950, surgiu a Teoria do Controle Ótimo criada por Lev Pontryagin⁴⁵ (1908-1988) e seus colaboradores que é uma generalização do Cálculo Variacional e foi rapidamente incorporada a Teoria Econômica. Portanto, a Dinâmica Econômica também está na sequência das análises do presente livro.

Esperamos que a provável satisfação do percurso inicial seja suficiente para motivar nossos leitores a continuarem seus estudos em outras obras e nos trabalhos originais de matemáticos e economistas.

¹ DALY, H.; FARLEY, J. *Economia Ecológica: Princípios e Aplicações*. Lisboa: Instituto Piaget, 2008, p. 22.

² SAMUELSON, P. *Foundations of Economic Analysis*. New York: Atheneum, 1974.

³ SAMUELSON, P. *Fundamentos de Análise Econômica*. São Paulo: Abril Cultural, 1983, p. 9.

⁴ SILBERBERG, E. “How to do Comparative Statics on the Back of an Envelope”. *Journal of Economic Theory*, v. 7, nº 2 (1974), pp. 159-172.

⁵ GALE, D., NIKAIÐÔ, H. “The Jacobian Matrix and Global Univalence of Mappings”. *Mathematische Annalen*, v. 159 (1965), pp. 81-93.

⁶ SAMUELSON, P. *Foundations of Economic Analysis*. New York: Atheneum, 1974, p. 257.

⁷ SAMUELSON, P. *Fundamentos de Análise Econômica*. São Paulo, Abril Cultural, 1983, p. 219 [capítulo IX].

⁸ Ibidem, pp. 217

⁹ CORNES, R. *Duality and Modern Economics*. New York: Cambridge University Press, 1992.

- ¹⁰ EICHHORN, W.; OETTLI, W. A General Formulation of the LeChatelier-Samuelson Principle. *Econometrica*, v. 40, n° 4 (1972).
- ¹¹ SCHMIDT, T. "Really Pushing the Envelope: Early Use of the Envelope Theorem by Auspitz and Lieben". *History of Political Economy*, v. 36, n° 1, pp. 103-129.
- ¹² SILBERBERG, E. "The Le Chatelier Principle as a Corollary to a Generalized Envelope Theorem". *Journal of Economic Theory*, v. 3, n° 2 (1971), pp. 146-155.
- ¹³ SILBERBERG, E. "How to Do Comparative Statics on the Back of an Envelope". *Journal of Economic Theory*, v. 7, n° 2 (1974), pp. 159-172.
- ¹⁴ HOTELLING, H. "Edgeworth's Taxation Paradox and the Nature of Supply and Demand Functions". *Journal of Political Economy*, v. 40, n° 5 (1932), p. 597.
- ¹⁵ SHEPHARD, R. *Theory of Cost and Production Functions*. Princeton: Princeton University Press, 1970, p. 171.
- ¹⁶ ROY, R. "La Distribution du Revenu entre les Divers Biens". *Econometrica*, v. 15, n° 3 (1947), p. 220.
- ¹⁷ SILBERBERG, E. "How to Do Comparative Statics on the Back of an Envelope". *Journal of Economic Theory*, v. 7, n° 2 (1974), pp. 159-172.
- ¹⁸ SLUTSKY, E. "Sulla Teoria del Bilancio del Consumatore". *Giornale degli Economisti e Rivista di Statistica*, v. 51, n°1 (1915), pp. 1-26.
- ¹⁹ SILBERBERG, E. "How to Do Comparative Statics on the Back of an Envelope". *Journal of Economic Theory*, v. 7, n° 2 (1974), pp. 159-172.
- ²⁰ SAMUELSON, P. *Foundations of Economic Analysis*. New York: Atheneum, 1974, p. 34.
- ²¹ SAMUELSON, P. *Fundamentos de Análise Econômica*. São Paulo: Abril Cultural, 1983, p. 37.
- ²² MIRSKY, L. *An Introduction to Linear Algebra*. Londres: Oxford University Press, 1955, p. 25.
- ²³ AITKEN, A. C. *Determinants and Matrices*. Edimburgo: Oliver and Boyd LTD., 1956, p. 98.
- ²⁴ SILBERBERG, E. "The Le Chatelier Principle as Corollary to a Generalized Envelope Theorem." *Journal of Economic Theory*, v. 3, n° 2 (1971), p. 149.
- ²⁵ LADY, G.; QUIRK, J. "The Scope of the LeChatelier Principle". *PHYSICA A-Statistical Mechanics and Its Applications*, v. 381, p 359.
- ²⁶ SAMUELSON, P. *Fundamentos de Análise Econômica*. São Paulo, Abril Cultural, 1983, p. 39.
- ²⁷ SAMUELSON, P. "An Extension of The Le Chatelier Principle". *Econometrica*, v. 28, n° 2 (1960), p. 370.
- ²⁸ SAMUELSON, P. *Fundamentos de Análise Econômica*. São Paulo, Abril Cultural, 1983, p. 39.
- ²⁹ SILBERBERG, E. "The Le Chatelier Principle as Corollary to a Generalized Envelope Theorem". *Journal of Economic Theory*, v. 3, n° 2 (1971), p. 149.
- ³⁰ SILBERBERG, E. "How to Do Comparative Statics on the Back of an Envelope". *Journal of Economic Theory*, v. 7, n° 2 (1974), p. 164.
- ³¹ NEUMANN, J. "A Model of General Economic Equilibrium". **The Review of Economic Studies**, v. 13, n° 1 (1945-1946), pp. 1-9.

- ³² KAKUTANI, S. A Generalizaion of Brouwer's Fixed Point Theorem. **Duke Mathematical Journal**, v. 8, n. 3, 1941, pp. 457-459
- ³³ NASH, J. F. Equilibrium Points in N-Person Games. **Proceedings of the National Academy of Sciences**, v. 36, 1950, pp. 48-49
- ³⁴ MCKENZIE, LIONEL. W. On Equilibrium in Graham's Model of World Trade and Other Competitive Systems. **Econometrica**, v. 22, n. 2, 1954, pp. 147-161
- ³⁵ ARROW, K. J., DEBREU, G. Existence of an Equilibrium for a Competitive Economy. **Econometrica**, v. 22, n. 3, 1954, pp. 265-290
- ³⁶ LEONTIEF, V. Quantitative Input and Output Relations in Economic Systems of the United States. **The Review of Economics and Statistics**, v. 18, n.3, 1936, pp. 105-125
- ³⁷ LEONTIEF, V. Structural Matrices of National Economies. **Econometrica**, v. 17, sup., 1949, p. 273-282.
- ³⁸ SRAFFA, P. Produção de Mercadorias por Meio de Mercadorias. Coleção Os Pensadores. SP: Abril Cultural, 1980
- ³⁹ HAWKINS,D.,SIMON, H. Note: Some Conditions of Macroeconomic Stability. **Econometrica**, v. 17, n. 3/4,1949, p. 245-248.
- ⁴⁰ DEBREU, G. Nonnegative Square Matrices. **Econometrica**, v. 21, n. 4, 1953, p. 597-609.
- ⁴¹ MORISHIMA, M. Marx in the Light of Modern Economic Theory. **Econometrica**, v. 42, n. 4, p. 611-32, 1974.
- ⁴² KURZ, H., SALVADORI, N. **Theory of Production, A Long-Period Analysis**. UH: Cambridge University Press, 1997
- ⁴³ SCHULTZ, H. **The Theory and Measurement of Demand**. Chicago: Chicago University Press, 1938.
- ⁴⁴ EVANS, G. C. A Simple Theory of Economic Crises. **Journal of American Statistical Association**, v. 26, n. 173, 1931, pp. 61-68
- ⁴⁵ PONTRYAGIN, L. S. et alli. **The Mathematical Theory of Optimal Processes**. NY: John Wiley & Sons, 1962

Lição ZERO

Apêndice – Equações Lineares

Sumário

Lição 0 – Apêndice – Equações Lineares

A.1 Objetivo

A.2 Os espaços vetoriais

A.3 Produto escalar

A.3.1 Lei dos cossenos

A.3.2 O Produto escalar

A.4 Planos e retas

A.5 Exercícios

A.6 Retas e planos na Economia

Lição 0

Apêndice – Equações Lineares

A.1 Objetivo

O objetivo desta lição é oferecer um ponto de partida. Alguns conceitos importantes para a leitura das lições subsequentes estão aqui sumarizados. A lição serve como uma breve revisão dos conteúdos que são imprescindíveis para a continuidade da leitura. Nós pressupomos que o leitor tenha algum conhecimento básico de Cálculo e Álgebra Linear.

Os problemas que nos interessam estão relacionados à minimização ou à maximização de funções $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e a apresentação de qualquer conteúdo está direcionada para tais problemas.

Uma igualdade linear é uma expressão como $c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n = d$ no espaço $\mathbb{R}^n$. Algebricamente, a equação acima corresponde ao que chamamos um conjunto afim, ou espaço afim, ou ainda a uma variedade linear, de dimensão $n-1$ nesse espaço. Uma interseção de m espaços afins, como acima, corresponde às soluções de um sistema linear com m equações lineares no $\mathbb{R}^n$ em $n-1$ incógnitas. Nossa preocupação a seguir é com esses conjuntos, suas dimensões e propriedades geométricas.

A partir da próxima lição, enfatizaremos os conjuntos definidos por sistemas de desigualdades lineares $c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n \leq d$, chamados poliedros ou simplexos. Depois, analisaremos conjuntos definidos

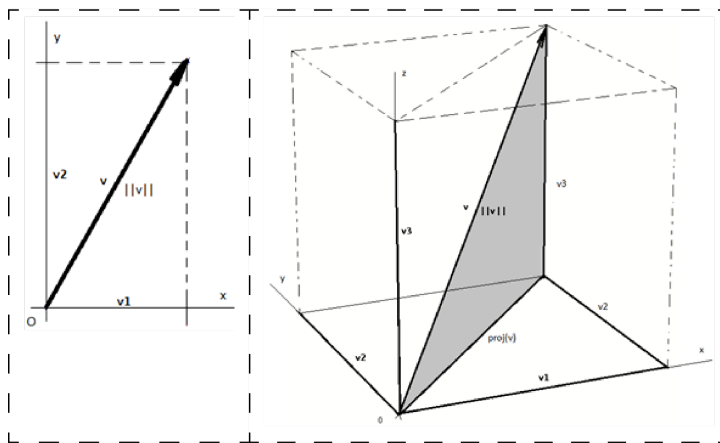
por sistemas gerais de equações $g(x_1, \dots, x_n) = d$ ou inequações $g(x_1, \dots, x_n) \leq d$. Num certo sentido, nosso curso é uma análise progressivamente mais complicada de um aspecto de sistemas de inequações.

A.2 Os espaços vetoriais $\mathbb{R}^n$

Os únicos espaços vetoriais que empregaremos são os conhecidos

$\mathbb{R}^n = \left\{ \vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mid x_i \in \mathbb{R} \right\}$. Para um vetor, usaremos a notação $\vec{v}$ ou $\vec{x}, \vec{y}$ etc. Usamos a notação $\vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ do $\mathbb{R}^n$ para diferenciar dos pontos $P(x_1, x_2, \dots, x_n)$ do espaço de coordenadas do espaço cartesiano $\mathbb{R}^n$. Entretanto, por comodidade gráfica, não manteremos essa notação, exceto por ênfase, adotando a mesma notação, $x = (x_1, x_2, \dots, x_n)$, tanto para ponto quanto para vetor. O contexto deve deixar evidente se estamos tratando de um ou outro.

A **norma ou comprimento** de um segmento orientado ou de um vetor no $\mathbb{R}^2$ ou $\mathbb{R}^3$ é o comprimento de qualquer segmento orientado que representa o segmento ou o vetor. Veja os desenhos abaixo:



A expressão algébrica da norma é uma consequência trivial do teorema de Pitágoras. No gráfico da esquerda, temos o segmento orientado $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$. Por Pitágoras, o comprimento desse segmento é

$\|v\| = \sqrt{v_1^2 + v_2^2}$. No desenho da direita, temos o segmento orientado $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \in \mathbb{R}^3$.

Observando que $(\text{proj}(v))^2 = v_1^2 + v_2^2$, resulta $\|v\| = \sqrt{v_1^2 + v_1^2 + v_2^2}$, novamente, por Pitágoras. Generalizando esses resultados para o $\mathbb{R}^n$, definimos a norma de um vetor $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ como $\|v\| = \sqrt{v_1^2 + \dots + v_n^2}$.

Um vetor retém, de cada segmento orientado finito do espaço cartesiano $\mathbb{R}^n$, o comprimento, a direção da reta sobre a qual está o segmento e, nessa reta, um sentido de percurso. O vetor identifica todos os segmentos que têm as mesmas características em relação ao comprimento, à direção e ao sentido de percurso, desconsiderando os pontos iniciais de cada segmento em particular. Por exemplo, quando dizemos que a frente fria se deslocou 500km na direção e sentido norte-sul, estamos falando de um deslocamento de 500km (comprimento), direção dada pela reta “norte-sul” percorrida no sentido do norte para o sul. No entanto, não se esclarece onde isso ocorreu. O ponto inicial pode estar em Porto Alegre, em Florianópolis, em Campo Grande ou em qualquer outra cidade. Deslocamento é um exemplo de vetor. Se, por exemplo, dizemos que o ponto inicial está em Florianópolis, temos um segmento orientado cuja extremidade final está próxima de Porto Alegre. Se dizemos que a extremidade inicial está em Porto Alegre, as pessoas que estão próximas à fronteira com o Uruguai devem se preocupar.

Conforme comentamos, em geral, não seguiremos rigorosamente as notações para segmentos orientados, vetores e pontos. Confundiremos as notações desde que o contexto permita claramente o reconhecimento do que é segmento orientado, vetor ou ponto. Distinguiremos quando suspeitarmos que o uso indistinto da notação possa resultar em confusão ou quando, por alguma razão, quisermos enfatizar as diferenças.

Se $\vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, e entendermos $\vec{v}$ como uma matriz $n \times 1$, então $\vec{v}' = (x_1 \ x_2 \ \dots \ x_n)$ é a matriz transposta – também usaremos a notação matricial.

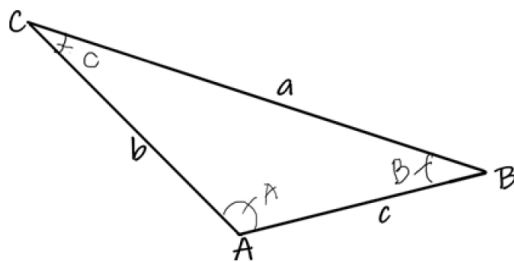
Um subespaço vetorial do $\mathbb{R}^n$ é qualquer subconjunto de $V \subseteq \mathbb{R}^n$ tal que se $v, w \in V$, então $v + w \in V$; se $\lambda \in \mathbb{R}, v \in V$, então $\lambda \cdot v \in V$ e $(0, 0, \dots, 0) = \vec{0} \in V$. Um subespaço afim $W \subseteq \mathbb{R}^n$ é $W = \{v + a \mid v \in V\}$, em que $V \subseteq \mathbb{R}^n$, é um espaço do $\mathbb{R}^n$.

A.3 Produto escalar

O conceito fundamental utilizado neste curso é o de **produto escalar** entre vetores do $\mathbb{R}^n$. Esse conceito resulta, muito naturalmente, da chamada “**Lei dos Cossenos**” que é, por sua vez, uma generalização do Teorema de Pitágoras.

A.3.1 Lei dos cossenos

Vejamos a lei dos cossenos. Seja o triângulo Δ como no gráfico abaixo:



A **lei dos cossenos** diz que o comprimento ao quadrado do lado a do triângulo é a soma dos quadrados dos comprimentos dos demais lados b e c menos o fator $2 \cdot c \cdot b \cdot \cos A$, em que $\cos A$ é o cosseno do ângulo no vértice A , oposto ao lado a :

$$a^2 = b^2 + c^2 - 2 \cdot c \cdot b \cdot \cos A.$$

No caso representado no desenho $0^\circ < A < 90^\circ$, mas a lei vale para qualquer ângulo $0^\circ \leq A \leq 180^\circ$. No caso extremo $A = 180^\circ$, temos $a = b + c$, que pela fórmula resulta em $a^2 = b^2 + c^2 + 2 \cdot c \cdot b$. No caso $0^\circ = A$, temos $a = b - c$, que pela fórmula resulta em $a^2 = b^2 + c^2 - 2 \cdot c \cdot b$. O fato que o ângulo no vértice A pode ser $0^\circ = A$ ou $A = 180^\circ$ significa, no caso $0^\circ = A$, que estamos descontando c de b , ou vice-versa, ou seja, $a^2 = (b - c)^2 = (c - b)^2 = b^2 + c^2 - 2 \cdot c \cdot b$, querendo dizer com isso que um dos lados, c ou b , colapsou sobre o outro e que a resultou como uma diferença entre c de b . Se $A = 180^\circ$ significa que estamos somando c de b , ou seja, $a^2 = (b + c)^2 = b^2 + c^2 + 2 \cdot c \cdot b$, querendo dizer com isso que os dois lados, c e b , estão sobre a mesma reta, mas em direções opostas e, nesse caso, a resulta na soma de c e b .

A lei dos cossenos é uma generalização do Teorema de Pitágoras – formulada para triângulos retângulos – para triângulos em geral. Adiante, na Lição Sete, veremos mais uma generalização do Teorema de Pitágoras, no caso, para matrizes, conhecida como “Lei de Binet-Cauchy”.

A.3.2 O Produto escalar

Observe o desenho abaixo. Sejam dois vetores, $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$, $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, do $\mathbb{R}^3$ e θ o ângulo entre eles:

Pela lei dos cossenos, $\|u - v\|^2 = u^2 + v^2 - 2 \cdot \|u\| \cdot \|v\| \cdot \cos \theta$, portanto,

$$(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2 = u_1^2 + u_2^2 + u_3^2 + v_1^2 + v_2^2 + v_3^2 - 2 \cdot \|u\| \cdot \|v\| \cdot \cos \theta.$$

Cancelando os termos u_i^2 e v_i^2 na esquerda e na direita da expressão acima, resulta

$$u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3 = \|u\| \cdot \|v\| \cdot \cos \theta.$$

O lado esquerdo da expressão acima,

$$u \cdot v = u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3,$$

é dito o **produto escalar** dos vetores u e v . Portanto, temos a expressão para o produto escalar

$$u \cdot v = \|u\| \cdot \|v\| \cdot \cos \theta, \text{ para } 0^\circ \leq \theta \leq 180^\circ.$$

Também notamos o produto escalar por $\langle u, v \rangle$, isto é, $\langle u, v \rangle = u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3$. Na notação matricial, escreveremos:

$$\vec{u} \cdot \vec{v} = (u_1 \quad u_2 \quad u_3) \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = u' \cdot v.$$

Obviamente, se dois vetores não nulos são ortogonais, isto é, $\theta = 90^\circ$, resulta $u \cdot v = 0$. Por essa razão, dizemos que dois vetores são **ortogonais** se $u \cdot v = 0$. Pode ocorrer que $\|u\| = 0$ ou $\|v\| = 0$, ou seja, um dos vetores ou ambos serem nulos, nesse caso não tem sentido definir ângulo entre os vetores e dizer que $\theta = 90^\circ$. Entretanto, por uma questão de conveniência, definimos a ortogonalidade entre dois vetores como $u \cdot v = 0$ com a observação de que o vetor nulo é ortogonal a todo vetor, inclusive a si mesmo. Mas, se ambos os vetores não são nulos, então $\theta = 90^\circ$ mantém o sentido geométrico. Observe que, caso $u \cdot v = 0$ para todo o vetor v , então $u = 0$, ou seja, um vetor que é ortogonal a todo vetor é nulo, pois se fizermos $v = u$, $\vec{u} \cdot \vec{v} = (u_1 \quad u_2 \quad u_3) \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = u' \cdot u = u_1^2 + u_2^2 + u_3^2 = 0$, o que implica $u_1 = u_2 = u_3 = 0$.

Como é usual, estendemos as observações acima para os vetores no $\mathbb{R}^n$. Se $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$, $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$, então o produto escalar entre esses vetores é definido como $u \cdot v = u_1 \cdot v_1 + \dots + u_n \cdot v_n$.

Como dois vetores sempre estão num plano, a análise que resulta em $u \cdot v = \|u\| \cdot \|v\| \cdot \cos \theta$, para $0^\circ \leq \theta \leq 180^\circ$, como fizemos anteriormente em relação aos vetores do $\mathbb{R}^3$, e nada se modifica, exceto pelo número de variáveis.

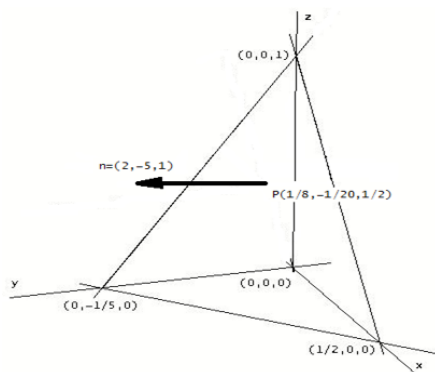
O conceito de ortogonalidade é definido de maneira idêntica: dois vetores são ortogonais se o produto escalar deles é nulo.

A expressão acima implica uma desigualdade que empregaremos futuramente: $|u \cdot v| \leq \|u\| \cdot \|v\|$, pois, obviamente $|u \cdot v| = \|u\| \cdot \|v\| \cdot |\cos \theta| \leq \|u\| \cdot \|v\|$.

A.4 Planos e retas

Sem dúvida, os primeiros exemplos de superfícies são aqueles representados por isoquantas de funções. Por exemplo, o plano $2 \cdot x - 5 \cdot y + z = 1$ é a isoquanta de nível 1 da função $F(x, y, z) = 2 \cdot x - 5 \cdot y + z$. Utilizemos esse caso especial para começar a nossa análise.

Primeiramente, o plano é constituído pelo conjunto dos pontos do $\mathbb{R}^3$ que satisfazem a condição $2 \cdot x - 5 \cdot y + z = 1$. Se nós isolamos a variável z , $z = 1 - 2 \cdot x + 5 \cdot y$, observamos que a superfície é descrita pelo gráfico $(x, y, 1 - 2 \cdot x + 5 \cdot y) \in \mathbb{R}^3$ da função $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = 1 - 2 \cdot x + 5 \cdot y$. Em princípio, podemos representar a mesma superfície tanto da maneira i) quanto da maneira ii) acima. Se descrevemos a superfície pelo gráfico de $f(x, y) = 1 - 2 \cdot x + 5 \cdot y$, substituindo $f(x, y)$ por z , obtemos a isoquanta $F(x, y, z) = 2 \cdot x - 5 \cdot y + z = 1$.



O plano acima é perpendicular ao vetor $\vec{n} = \begin{pmatrix} 2 \\ -5 \\ 1 \end{pmatrix}$. Pois, dados dois pontos sobre esse plano, digamos $P_0(x_0, y_0, z_0)$ e $P_1(x_1, y_1, z_1)$, então:

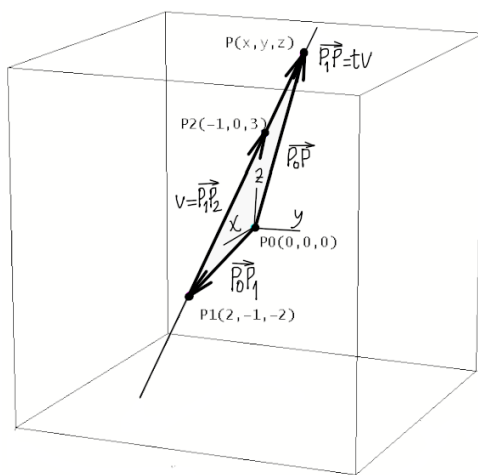
$$2 \cdot (x_1 - x_0) - 5 \cdot (y_1 - y_0) + (z_1 - z_0) = 0,$$

ou seja, o vetor $\vec{v} = \begin{pmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{pmatrix}$, representado pelo o segmento orientado $\vec{P_0P_1}$, é perpendicular ao vetor $\vec{n}$, pois $\vec{n} \cdot \vec{v} = 0$.

Vamos definir “planos” no $\mathbb{R}^n$. Dado um ponto $P^*(x_1^*, x_2^*, \dots, x_n^*) \in \mathbb{R}^n$ e um vetor $\vec{n} = (a_1^*, a_2^*, \dots, a_n^*)$, o **hiperplano de dimensão $(n-1)$** , que contém o ponto P^* e é perpendicular ao vetor normal $\vec{n}$, é definido como o conjunto dos pontos $P(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, tais que os vetores $\vec{v} = \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \\ \vdots \\ x_n - x_n^* \end{pmatrix}$ são ortogonais ao vetor $\vec{n}$, ou seja, $\vec{v} \cdot \vec{n} = 0$, ou ainda, $a_1^* \cdot (x_1 - x_1^*) + a_2^* \cdot (x_2 - x_2^*) + \dots + a_n^* \cdot (x_n - x_n^*) = 0$.

Antes de respondermos à questão se poderíamos representar uma reta no $\mathbb{R}^3$ também como uma isoquanta, vejamos um exemplo.

Exemplo 1: A reta que passa pelo ponto $P_0(2, -1, -2)$ e está na direção do vetor $\vec{v} = \begin{pmatrix} -3 \\ 1 \\ 5 \end{pmatrix}$ pode ser representada pelo conjunto de equações paramétricas $x = 2 + (-3) \cdot t$, $y = -1 + 1 \cdot t$, $z = -2 + 5 \cdot t$.



Como $\vec{v} = \vec{P_1P_2}$, dado um ponto qualquer $P(x, y, z)$, então $\vec{P_0P} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \vec{P_0P_1} + t \cdot \vec{v} = \vec{P_0P_1} + t \cdot \vec{P_1P_2}$, que é a equação vetorial da reta. Fazendo variar $t \in \mathbb{R}$ obtemos todos os pontos $P(x, y, z)$ sobre a reta.

Essas equações são ditas paramétricas porque, ao variar o parâmetro t , obtemos diferentes pontos sobre a reta. Essas equações

não são evidentemente únicas, pois, como essa reta passa também pelo ponto $P\left(\frac{4}{5}, -\frac{3}{5}, 0\right)$, faça $t = \frac{2}{5}$ nas equações anteriores e o vetor $\vec{u} = \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix}$ será paralelo ao vetor $\vec{v}$, então as equações paramétricas abaixo representam a mesma reta:

$$x = \frac{4}{5} - \frac{3}{4} \cdot s, \quad y = -\frac{3}{5} + \frac{1}{5} \cdot s, \quad z = s.$$

Variando o parâmetro s obtemos os pontos da reta.

Exemplo 2: Cada uma das equações abaixo representa um plano:

$$\pi_1: 2 \cdot x + y + z = 1,$$

$$\pi_2: x - 2 \cdot y + z = 2.$$

Observe que esses planos não são paralelos, pois os vetores $\vec{n}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ e $\vec{n}_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, que são perpendiculares aos planos π_1 e π_2 , respectivamente, não são paralelos. Logo, os pontos $P(x, y, z)$, que solucionam as duas equações simultaneamente, formam a reta que está na interseção desses planos. Por linha-redução é fácil obter essa reta, pois

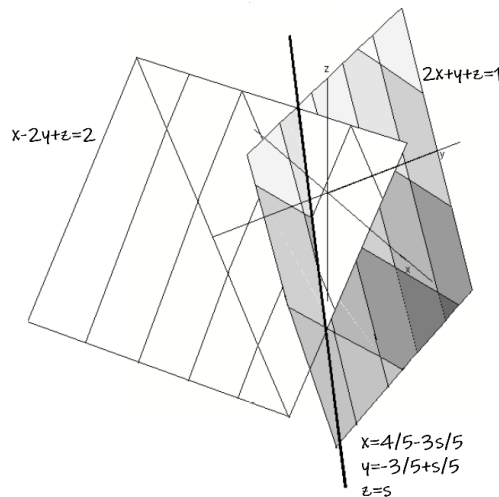
$$\left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 1 & -2 & 1 & 2 \end{array} \right) \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & 1 & -\frac{1}{5} & -\frac{3}{5} \end{array} \right),$$

ou seja, $x = \frac{4}{5} - \frac{3}{5} \cdot z$ e $y = -\frac{3}{5} + \frac{1}{5} \cdot z$. Se nós substituirmos a variável z por s e escrevemos

$$x = \frac{4}{5} - \frac{3}{5} \cdot s, \quad y = -\frac{3}{5} + \frac{1}{5} \cdot s \quad \text{e} \quad z = 0 + 1 \cdot s,$$

obtemos um conjunto de equações paramétricas para essa reta. Facilmente, vemos que a reta passa pelo ponto $P\left(\frac{4}{5}, -\frac{3}{5}, 0\right)$ na direção do

vetor $\vec{u} = \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix}$, ou seja, a reta do penúltimo exemplo acima. Portanto, podemos representar uma reta no $\mathbb{R}^3$ como a interseção dos dois planos acima: $2 \cdot x + y + z = 1$ e $x - 2 \cdot y + z = 2$.



Uma “reta” no $\mathbb{R}^n$ é definida como o conjunto de pontos do $\mathbb{R}^n$ determinado pelo ponto $P^*(x_1^*, x_2^*, \dots, x_n^*) \in \mathbb{R}^n$ e pelo vetor $\vec{v} = \begin{pmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_n^* \end{pmatrix}$, que dá a direção da reta, através das equações paramétricas abaixo:

$$x_1 = x_1^* + t \cdot b_1^*, x_2 = x_2^* + t \cdot b_2^*, \dots, x_n = x_n^* + t \cdot b_n^*, t \in \mathbb{R}.$$

A.5 Exercícios

- Determinar a equação do plano que contém o ponto $P(2, 0, 3)$ e é perpendicular ao vetor $\vec{n} = (-5, 2, -1)$.
- Determinar a equação do plano que contém o ponto $P(-2, 1, 5)$ e é perpendicular à reta que passa pelos pontos $P(1, -1, 2)$ e $P(2, -1, 3)$.
- Determinar a equação do plano que contém o ponto $P(-2, 1, 3)$ e é paralelo ao plano $2 \cdot x - y + z = 5$.

- iv) Determinar a equação do plano que contém o ponto $P(4,1,-5)$ e é perpendicular à reta $x = -3 + 5 \cdot t$, $y = 2 - t$, $z = 5 + t$.
- v) Determinar equações paramétricas para a reta que passa pelo ponto $P(-1,2,-5)$ e é paralela ao vetor $\vec{v} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$. Escolha outro ponto sobre a mesma reta e outro vetor na mesma direção da reta e escreva outras equações paramétricas descrevendo a reta anterior.
- vi) Determinar as equações paramétricas da reta que passa pelos pontos $P(-2,1,6)$ e $P(7,0,3)$.
- vii) Determinar as equações paramétricas da reta que passa pelo ponto $P(-1,0,5)$ e é perpendicular ao plano $3 \cdot x - y + 2 \cdot z = 6$.
- viii) Determinar um vetor na direção da reta:

$$\begin{aligned} x - y - z &= 0 \\ 2 \cdot x - y - 4 \cdot z &= -3. \end{aligned}$$

- ix) Determinar as equações paramétricas da reta na interseção dos planos:

$$3 \cdot x - 2 \cdot y + z = 0$$

$$2 \cdot x + y - 2 \cdot z = 1.$$

- x) Determinar um ponto pertencente ao hiperplano abaixo e um vetor perpendicular a este hiperplano:

$$x_1 + 2 \cdot x_2 - x_3 + x_4 - x_5 = 7.$$

- xi) Verifique que a interseção dos hiperplanos:

$$3 \cdot x_1 + 2 \cdot x_2 + x_3 - x_4 + 3 \cdot x_5 = 2$$

$$5 \cdot x_1 + 4 \cdot x_2 + 3 \cdot x_3 + 2 \cdot x_4 - x_5 = 3$$

Pode ser dada parametricamente como:

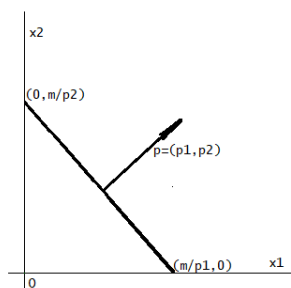
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1 \\ -1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t_1 \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t_2 \cdot \begin{pmatrix} 4 \\ -11/2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t_3 \cdot \begin{pmatrix} -7 \\ 9 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Observe que os vetores $\begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 4 \\ -11/2 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} -7 \\ 9 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ são linearmente independentes.

i) Determinar a equação do plano que contém os pontos $P(-2,1,5)$, $P(1,-1,2)$ e $P(2,-1,3)$. Use o exercício 1 sobre o produto vetorial para obter um vetor perpendicular ao plano.

A.6 Retas e planos na Economia

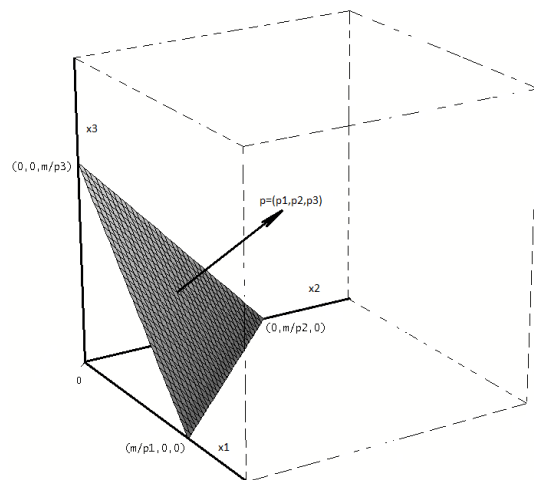
Retas e planos aparecem muito naturalmente na teoria econômica. Por exemplo, uma restrição orçamentária para uma cesta de dois bens matematicamente se expressa pelo segmento de reta $p_1 \cdot x_1 + p_2 \cdot x_2 = m$, $x_1 \geq 0$, $x_2 \geq 0$ unindo as extremidades $\left(\frac{m}{p_1}, 0\right)$ e $\left(0, \frac{m}{p_2}\right)$, em que m é a renda disponível para consumo destes bens e p_1 e p_2 são os preços. Observe que o vetor $p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ é ortogonal à reta $p_1 \cdot x_1 + p_2 \cdot x_2 = m$. Isso é muito importante:



De maneira similar, para uma cesta de três bens, uma restrição orçamentária é dada pelo simplexo $p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 = m$, $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$, que é simplesmente o triângulo com vértices $\left(\frac{m}{p_1}, 0, 0\right)$, $\left(0, \frac{m}{p_2}, 0\right)$ e $\left(0, 0, \frac{m}{p_3}\right)$.

Novamente, observe que o vetor $p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ é ortogonal ao plano

$$p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 = m.$$



É claro que podemos generalizar para uma cesta com n bens, considerando o simplexo $p_1 \cdot x_1 + p_2 \cdot x_2 + \dots + p_n \cdot x_n = m$, $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_n \geq 0$, que tem os vértices $\left(\frac{m}{p_1}, 0, \dots, 0\right)$, $\left(0, \frac{m}{p_2}, 0, \dots, 0\right)$, ..., $\left(0, \dots, 0, \frac{m}{p_n}\right)$. Novamente, observe que o vetor $p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$ é ortogonal ao hiperplano $p_1 \cdot x_1 + p_2 \cdot x_2 + \dots + p_n \cdot x_n = m$.

Muito importante é o chamado **simplexo** básico de dimensão $n-1$, $x_1 + x_2 + \dots + x_n = 1$, $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_n \geq 0$, cujos vértices são $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, \dots, 0)$, ..., $e_n = (0, 0, \dots, 1)$, que desempenha um papel fundamental em Teoria dos Jogos.